DOCUMENT RESUME

ED 448 225 UD 033 882

| | |
|---|---|
| AUTHOR | Molnar, Alex |
| TITLE | Smaller Classes Not Vouchers Increase Student Achievement. |
| INSTITUTION | Keystone Research Center, Harrisburg, PA. |
| PUB DATE | 1998-06-00 |
| NOTE | 54p. |
| AVAILABLE FROM | Keystone Research Center, 412 North Third Street, Harrisburg, PA 17101 ($12). Tel: 717-255-7181; Fax: 717-255-7193; e-mail: KeystoneRC@aol.com. |
| PUB TYPE | Opinion Papers (120) |
| EDRS PRICE | MF01/PC03 Plus Postage. |
| DESCRIPTORS | *Academic Achievement; *Class Size; *Educational Vouchers; Elementary Secondary Education; High Risk Students; Poverty; Private Schools; Public Schools; School Choice; *Small Classes |
| IDENTIFIERS | California; Cleveland Public Schools OH; Indiana; Milwaukee Public Schools WI; Nevada; Pennsylvania; Tennessee |

ABSTRACT

This report is designed to help Pennsylvania policymakers compare two current school reform ideas that are intended to improve student achievement: private school vouchers and class size reduction. It begins with a brief history of educational vouchers, then reviews the achievement effects of the Milwaukee, Wisconsin, and Cleveland, Ohio, educational voucher programs. The reported benefits of these programs are compared to the benefits of reducing class size in grades K-3, using data from studies on class size conducted in Tennessee, Indiana, Nevada, California, and Wisconsin. The report concludes with policy recommendations for Pennsylvania based on evidence which indicates that small classes generate the greatest gains in kindergarten and grade 1. These recommendations include providing universal, publicly funded, full-day kindergarten with student-teacher ratios of 15:1; reducing class size in first grade to 15 students; and implementing an experimental program of class size reductions in grades 2 and 3 in which class size reductions are achieved in a variety of ways. (Contains 142 references.) (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

KEYSTONE
RESEARCH
CENTER

ED 448 225

# Smaller Classes Not Vouchers Increase Student Achievement

Alex Molnar

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
C. Ramsey
Keystone Research Center
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

BEST COPY AVAILABLE





UD033882

# Smaller Classes

# Not Vouchers

# Increase Student

# Achievement

Alex Molnar

For additional copies of this report contact:

Keystone Research Center • 412 North Third Street • Harrisburg, PA 17101
Phone: 717/255-7181 • Fax: 717/255-7193 • Internet: KeystoneRC@aol.com

KEYSTONE RESEARCH CENTER

## The Keystone Research Center

The Keystone Research Center (KRC), a non-partisan think tank based in Harrisburg, conducts research on the Pennsylvania economy and labor market.

## Acknowledgments

Barbara Nye, Charles Achilles, and Alan Krueger assisted with interpreting the main findings of research on Tennessee's experience with reducing class size. Cecilia Rouse of Princeton generously reviewed the discussion of the Milwaukee voucher program. Brad Adams of the Wisconsin Department of Public Instruction and Ken Black, Senior Budget Analyst for the Milwaukee Public Schools, provided valuable information and insights about the Milwaukee voucher programs. Bert Holt of the Cleveland Scholarship and Tutoring Program, and Francis Rogers of the Ohio Department of Education did the same on the Cleveland voucher program. Mary Snow of the Nevada Department of Education supplied insights and sources on Nevada's class size reduction initiative. Mary Beth Morgan reviewed the discussion of Indiana's Prime Time program. Joan McRobbie of WestEd and Professor Michael Kirst of Stanford provided the latest sources and perspective on California's class-size reduction program. Howard Wial of Keystone contributed to the discussion of vouchers and equity and reviewed the entire manuscript. Graphic artist Lori Baker of *The Design Department* helped make the document readable and showed unending good cheer through several rounds of edits. Finally, this project would not have come to fruition without the tireless and deft editorial work of Stephen Herzenberg.

## Author

Alex Molnar holds a Ph.D. in Urban Education and has been a Professor in the School of Education at the University of Wisconsin-Milwaukee since 1972. Previously, he taught high-school social studies in the Chicago area. From 1993-95, Professor Molnar served as chief of staff for the Wisconsin Department of Public Instruction Urban Initiative. He is currently a member of the legislatively-mandated evaluation team for the Wisconsin small schools initiative—Student Achievement Guarantee in Education (SAGE). Professor Molnar has edited, written, or co-authored several books, including *Changing Problem Behavior in Schools* (1989), which is widely used by American educators and has been translated into four languages. Professor Molnar consults extensively with school districts throughout the United States.

© Keystone Research Center, 1998

KEYSTONE RESEARCH CENTER

# TABLE OF CONTENTS

EXECUTIVE SUMMARY . . . . . 1

INTRODUCTION . . . . . 6

VOUCHERS . . . . . 7
- Historical Background . . . . . 7
  - Educational Choice Enters The Mainstream . . . . . 8
  - The Battle Over Vouchers Today . . . . . 9

- The Milwaukee Parental Choice Voucher Program . . . . . 10

- The Debate Over The Achievement Effect Of The Milwaukee Voucher Program . . . . . 12
  - Why Different Researchers Reach Different Conclusions . . . . . 15
    - The Witte Evaluations . . . . . 16
    - The Greene, Peterson, and Du Evaluation . . . . . 17
    - The Rouse Evaluation . . . . . 20
  - Milwaukee's Private Voucher Program—PAVE . . . . . 22

- The Cleveland Scholarship And Tutoring Program (CSTP) . . . . . 23

- Vouchers, Values, And Educational Equity . . . . . 25

SMALL CLASS SIZES . . . . . 28
- Historical Background . . . . . 28
  - Indiana Prime Time . . . . . 29

- The Tennessee Experience . . . . . 30
  - The Tennessee Star Study . . . . . 30
  - The Tennessee Lasting Benefits (LBS) Study . . . . . 33
  - Project Challenge . . . . . 33

- Nevada and California . . . . . 34
  - Nevada . . . . . 34
  - California . . . . . 34

- Wisconsin: Student Achievement Guarantee In Education (SAGE) . . . . . 36

RECOMMENDATIONS . . . . . 38

REFERENCES . . . . . 40

KEYSTONE RESEARCH CENTER

# EXECUTIVE SUMMARY

Fiscal constraints make it imperative that Pennsylvania allocate money to those school reform initiatives that hold out the greatest promise of success. The purpose of this report is to help policy makers compare two current school reform ideas: private school vouchers and class size reductions.

A brief history of vouchers precedes a review of the achievement effects of the Milwaukee and Cleveland voucher programs. The reported benefits of these programs are compared to the benefits of reducing class size in grades K-3. The report concludes with policy recommendations.

Each of the major sections of the report is largely self-contained. Readers can therefore go immediately to the sections that most interest them. Some readers, for example, may want to skip the historical background on vouchers and jump to the discussion of their achievement effects in Milwaukee. Others may want to begin by reading about class size.

---

**VOUCHERS:** Nearly all of the achievement evidence that exists on vouchers comes from the Milwaukee Parental Choice Program.

- Five legislatively mandated evaluations of the Milwaukee program found no achievement gains for voucher students.
- While two reexaminations of the same data have found an achievement advantage in math for voucher students, both have significant flaws that cast doubt on their findings.
  - The first reexamination compares voucher students to a small (26 students in one of the years) and unrepresentative group of students.
  - The second assumes that voucher students—despite their having more educated parents with higher academic expectations—would not have achieved more over time wherever they studied.
- The Milwaukee program has never enrolled more than 1,650 students in a given year.
  - Data limitations and small sample size plague any attempt to analyze the Milwaukee experience.
- A December 1997 analysis found that the best schools in Milwaukee are a group of 14 public schools with small classes that serve an economically disadvantaged population.
  - Students at these public schools match voucher students on math tests and outperform them on reading tests.
- *In sum, no strong evidence exists that participation in a voucher program increases student achievement.*

KEYSTONE RESEARCH CENTER

**SMALL CLASS SIZE:** Conclusive evidence from multiple sources shows that reducing class size improves student achievement.

**The Tennessee Experience:** The Tennessee Student/Teacher Achievement Ratio (STAR) project was the single most definitive class size study. Beginning in 1985, the state Department of Education randomly assigned kindergarten and first grade students to small classes (about 13-17 students), regular classes (about 22-25 students), and regular classes with an instructional aide. Once assigned to small classes, students remained in them.

- On average, students attending small classes in K–3 achieved scores substantially higher than students in regular classes.
- Increasing the number of teachers' aides had only a very small positive impact on test scores.
- Achievement gains from small size Tennessee classes have lasted through at least 8th grade.
- Lower achieving, minority, and poor students benefited most from small classes in Tennessee.

Figures 1 and 2 illustrate how much Tennessee students in small classes outperformed other students in every geographical setting. Small class students in inner-city areas enjoyed the biggest achievement gains.

**Figure 1: Third Grade Math Achievement in Tennessee—the Impact of Small K–3 Classes and Aides**

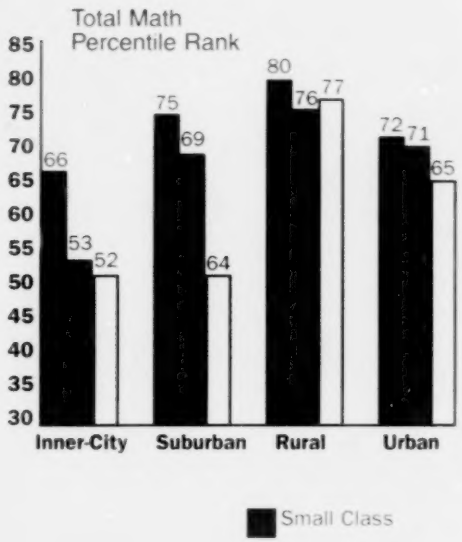
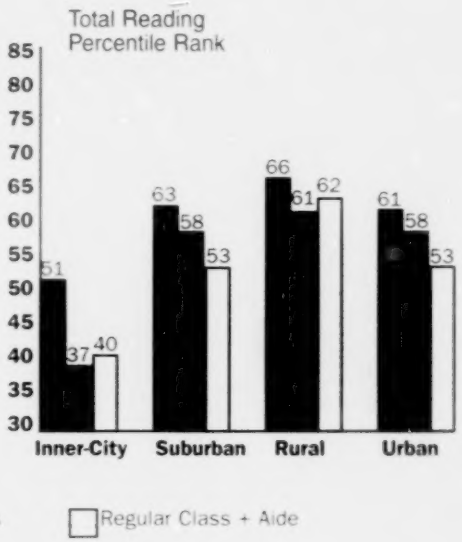



**Figure 2: Third Grade Reading Achievement in Tennessee—the Impact of Small K–3 Classes and Aides**




Small Class  Regular Class  Regular Class + Aide

Note: For additional discussion of the research from which these results came, see the section of this report on the Tennessee STAR Project.

Source: Elizabeth Word, Charles M. Achilles, Helen Bain, John Folger, John Johnston and Nan Lintz, "Project STAR Final Executive Summary Report" (Nashville, TN: Tennessee State Department of Education, June 1990).

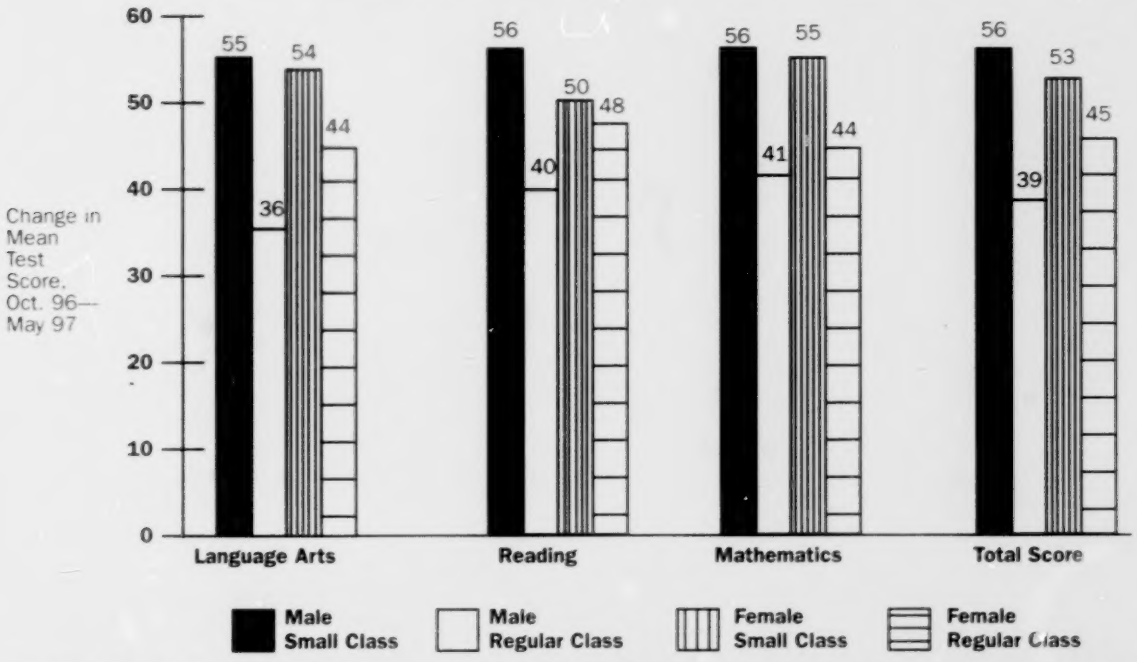
**The Wisconsin Experience:** In December 1997, in preliminary findings from Wisconsin's Student Achievement Guarantee in Education (SAGE) class-size initiative, Alex Molnar and co-authors reported results consistent with Tennessee experience. The SAGE initiative is the largest, most systematic study of class size since Project STAR. In its first year, SAGE involved 3600 students and targeted small classes to low-income areas throughout the state.

- Between October 1996 and May 1997, the increase in test scores for first grade students in SAGE schools exceeded by 12–14 percent the increase in scores for students in a comparison group of schools with regular size classes.
- In SAGE classrooms, the total scores achieved by African-American males on three tests increased by over 40 percent more than African-American male scores in a comparison group of schools. (See Figure 3 below.)
- After controlling for individual differences among students (e.g., race, subsidized lunch eligibility, days absent), SAGE students enjoyed significantly greater improvements in test scores in reading, language arts, and math.

**National Evidence:** A study by Harold Wenglinsky (Educational Testing Service) of math achievement in 203 school districts across the country gives an indication of the size of the cumulative benefits of small classes by fourth grade.

- Fourth graders in smaller-than-average classes were about four months ahead of fourth graders in larger-than-average classes.
- In the sub-group of schools that included mainly large urban areas, fourth graders in smaller-than-average classes were three-quarters of a school year ahead of their counterparts in larger-than-average classes.

**Figure 3: Increase in African–American First Grade Test Scores in Low–Income Wisconsin Schools, Small vs. Regular Classes**




Source: Peter Maier, Alex Molnar, Philip Smith, John Zahorik, *First-year Results of the Student Achievement Guarantee in Education Program* (Milwaukee: Center for Urban Initiatives and Research, University of Wisconsin–Milwaukee, December 1997), Table 33.

KEYSTONE RESEARCH CENTER

**RECOMMENDATIONS:** To improve educational outcomes quickly and across the board, Pennsylvania should reduce class size in K–3.

The evidence indicates that small classes generate the greatest gains in kindergarten and first grade. The Keystone Research Center therefore recommends that Pennsylvania:

- Provide universal, publicly funded full-day kindergarten with student-teacher ratios of 15:1; and
- Reduce class size in first grade to 15.

Research suggests that more modest gains result from small classes in grades two and three. In addition, considerable scope for innovation exists in exploring how to build on gains established in small kindergarten and first grade classes. Therefore, the Keystone Research Center recommends that Pennsylvania:

- Implement an experimental program of class size reductions in grades two and three. This program should evaluate the effectiveness of achieving class size reductions in various ways (e.g., in the main instructional subjects only), and in combination with other (e.g., curricular and teacher training) innovations.

To make for a smooth transition and avoid teacher and classroom shortages, these recommendations should be phased in, starting with kindergarten in the first year and first grade in the second. Implementation should be targeted initially at the schools and communities most in need—those in the bottom quarter of schools, measured by income and test scores.

Small class sizes and all-day kindergarten should be implemented systematically, with researchers collaborating with policymakers and practitioners so that lessons learned in the early stages allow for cost-effective implementation of small classes for all K-3 students in the state.

To implement these recommendations would cost the state an estimated $100 million in each of the first two years. This is a small fraction of Pennsylvania's projected budget surplus for 1997-98. It amounts to an annual investment of about $8.33 by each for the state's 12 million residents.

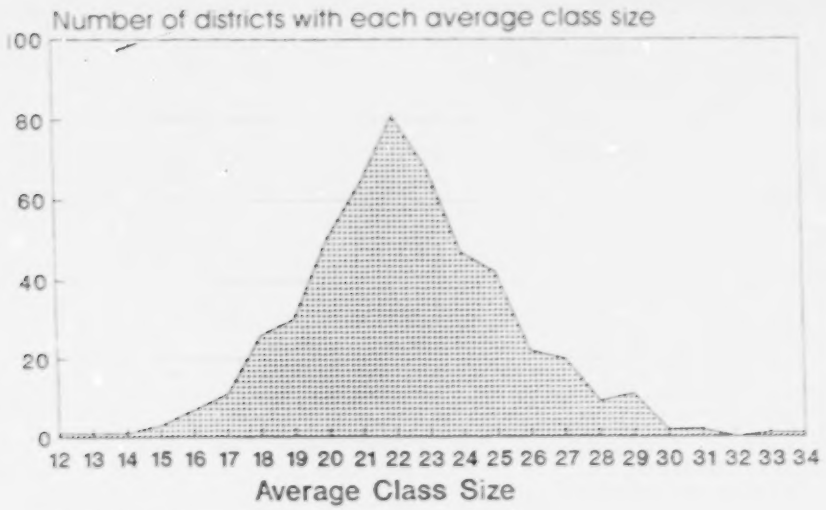
**BOX 1: CLASS SIZE IN PENNSYLVANIA TODAY**

Research shows that class sizes of under 20 lead to significant improvements in student performance. Only 20 percent of classes in Pennsylvania schools with elementary grades only are this small. Another 19 percent of classes in these schools have 27 or more students, about twice as big as the 15-student classes achieved in the Wisconsin and Tennessee class-size experiments. Figures 4 and 5 (on the next page) show the number of Pennsylvania school districts with kindergarten and first grade classes of different sizes in the late 1980s. Kindergarten classes usually had about 21–22 students and first grade classes about 23–24.

| | Percent of Classes in Each Size Range | | |
|---|---|---|---|
| **Average Class Size in 1995-96 (# of Students)** | **Schools with Elementary Grades Only** | **Schools with Secondary Grades Only** | **Schools with Elementary and Secondary Grades** |
| 1-20 | 20.5 percent | 28.8 percent | 11.0 percent |
| 21-23 | 29.9 | 20.5 | 14.4 |
| 24-26 | 30.3 | 23.1 | 21.0 |
| 27-29 | 14.4 | 16.4 | 30.0 |
| 30+ | 4.8 | 11.6 | 23.7 |

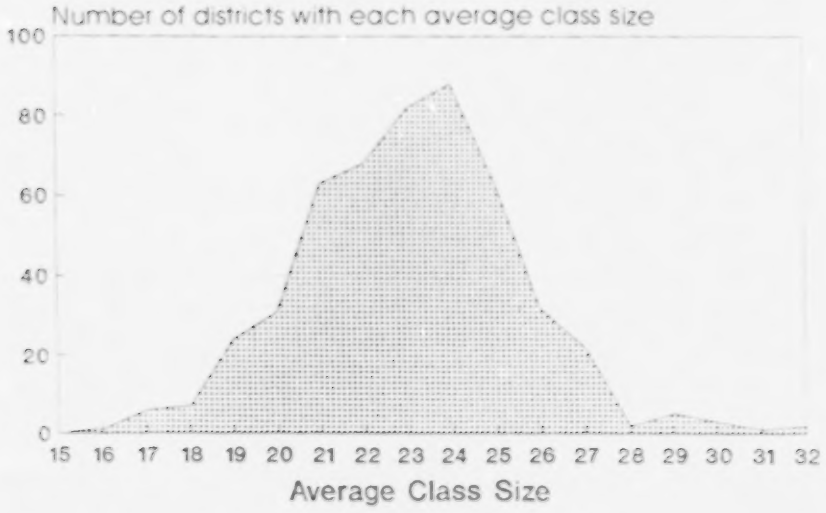
Source: Pennsylvania Department of Education, Pennsylvania System of School Assessment School Profiles, www.paprofiles.org.

## Figure 4: Kindergarten Class Sizes in Pennsylvania




Source: 1989 survey by office of Representative Ron Cowell

## Figure 5: First Grade Class Sizes in Pennsylvania




Source: 1989 survey by office of Representative Ron Cowell

KEYSTONE RESEARCH CENTER

# INTRODUCTION

Fiscal constraints make it imperative that Pennsylvania allocate money to those school reform initiatives that hold out the greatest promise of success. The purpose of this report is to help policy makers compare two current school reform ideas, private school vouchers and class size reductions.

A brief history of vouchers precedes a review of the achievement effects of the Milwaukee and Cleveland voucher programs. The reported benefits of these programs are compared to the benefits of reducing class size in grades K-3. The report concludes with policy recommendations.

Each of the major sections of the report is largely self-contained. Readers can therefore go immediately to the sections that most interest them. Some readers, for example, may know enough to skip the historical background on vouchers and jump to the discussion of their achievement effects in Milwaukee. Others may want to begin by reading about class size.

The research which this report reviews uses a variety of technical terms in analyzing the impact of vouchers and small classes. To make the report more comprehensible, the definitions of the key terms are listed in Box 2.

## BOX 2: WHAT ARE THESE RESEARCHERS TALKING ABOUT? A GLOSSARY OF TERMS

**CSTP:** The Cleveland Scholarship and Tutoring Program, the official name of the Cleveland voucher program.

**Control (as in "control for" and "control group"):** To evaluate the impact of a voucher program or smaller class size on student achievement, analysts need to isolate their impact from other variables (such as a students' family background). This can be done by comparing the performance of the students who get vouchers or attend small classes with the performance of another group of students—a "control group"—that is as similar as possible except for not having received vouchers or attended a small class. In addition, in statistical analysis, researchers usually take explicit account of—or "control for"—family and individual difference, so that the impact of vouchers or class size will not be incorrectly estimated.

**Effect size:** To evaluate the benefits of vouchers or smaller class sizes, you need to know how big an impact they have on student achievement. Effect sizes gauge this impact by looking at the gap in test scores between students who receive vouchers or attend small classes and the scores of students who don't. This gap is divided by a measure of the overall spread of student scores. (See standard deviation).

**Meta-analysis:** When a large number of studies have been conducted on a subject—such as the achievement impact of small class size—a systematic evaluation, or meta-analysis, of these prior studies may be used as a tool for determining the overall weight of the evidence. In weighing the importance of each study, the meta-analysis takes into account such factors as its sample size and the methodological rigor used.

**MPS:** Milwaukee Public Schools.

**MPCP:** Milwaukee Public Choice Program, the official name of the Milwaukee voucher program.

**Percentile ranks:** To evaluate the benefits of vouchers or smaller class sizes, you need to know how big an impact they have on student achievement. One way to do this is by considering how much an improvement in test scores would have moved a student up in the overall student ranking. If an improvement would move a student up from, say, the mid-point of the achievement curve (the 50th percentile) past another 10 percent of students (to the 60th percentile), it would be said to have improved scores by 10 percentile ranks.

**Standard deviation** is a measure of how spread out a group of numbers (such as student test scores) is. It equals the square root of the average squared difference between test scores and the average test score.

**Statistical significance:** In evaluating the impact of vouchers or class size on test scores (or of any variable on another variable), researchers want to know whether they can be confident that an observed performance difference is large enough that it could not have occurred by random chance. If the difference is so large that it could only have occurred by chance with a small probability ("small" being defined customarily as 5 times out of 100), then the observed change in performance is considered to be statistically significant.

KEYSTONE RESEARCH CENTER

# VOUCHERS

## Historical Background

In the early 1870s, demoralized by their crushing defeat in the Franco-Prussian War, many French citizens angrily blamed the public school system for their woes. They declared that it was "the Prussian teacher [who] has won the war."[1]

To improve the schools, and presumably France's prospects in the next war, a French parliamentary commission in 1872 recommended a religious school voucher plan remarkably similar to the ones currently being proposed in the United States. In 19th century France, however, hostility to the idea of providing public money to church schools was so widespread that the French Assembly never took up the plan.

Just over 100 years later, with the U.S. trade deficit at record levels, the authors of *A Nation at Risk* declared that America was headed for a disastrous defeat in a global economic war.[2] As in nineteenth-century France, the public schools were called to account. *A Nation at Risk* helped make the belief that the U.S. system of public education is a catastrophic failure an article of faith in the nation's school reform deliberations. In so doing it helped set the stage for school voucher proposals in the late 1980s and 1990s.

Until the 1980s, the constitutional prohibition against church-state entanglements, public opposition to the use of tax funds for religious schools, and a lack of a generally available alternatives to public schools kept voucher proposals on the fringes of American school reform.

Educational vouchers were first proposed in the United States in 1955 by economist Milton Friedman.[3] Friedman argued for providing parents with vouchers and allowing them to choose any school, public or private, for their children to attend. In his view, an educational market would be more efficient at allocating educational resources than a system of government-run schools. Friedman's idea initially drew scant attention and little support.

The private school choice plans proposed in the United States in the late 1950s and early 1960s were not motivated by a desire to create competition and an educational market. These plans grew out of opposition to court-ordered desegregation in the wake of the 1954 U.S. Supreme Court's *Brown v. Board of Education decision.*[4] The Virginia legislature in 1956 passed a "tuition-grant" program and in 1960 a "scholarship" plan that provided students with tax dollars to pay the tuition at any qualified non-sectarian school in their district. The Virginia laws and other "freedom of choice" plans passed by southern legislatures expressly sought to help maintain segregated school systems.

Since the late 1950s, private school choice has moved into the mainstream school reform debate. Private school vouchers have found support among three groups:

1) Catholics who see taxpayer-financed vouchers as a fiscal life line for their cash poor schools (some Catholics remained opposed to vouchers because they feared that public funding would increase public regulation of religious schools);

2) Free-market advocates who regard vouchers as a way of increasing efficiency in the provision of public education;

3) People of all political persuasions who, for various reasons, are dissatisfied with the shortcomings of what David Tyack, an historian of public education, has labeled "the one best system."[5]

In the late 1960s, the Democratic administration of President Lyndon Johnson embraced the idea of vouchers. At the time, the voucher constituency included not only some political conservatives and segments of the business community, but also "de-schoolers" influenced by the writing of Ivan Illich,[6] progressive and black nationalist "free schoolers,"[7] social critics of the public education bureaucracy such as Paul Goodman,[8] and liberal academics like Christopher Jencks.[9] The chance to craft "regulated" voucher plans—ensuring that the poorest recipients got the largest vouchers—appealed to many liberals.

The administration of President Richard Nixon subsequently advanced the Johnson proposal. However, little local enthusiasm emerged for the idea. Minneapolis, Rochester, Kansas City, Milwaukee, Gary, and Seattle all rejected the opportunity to participate. Only Alum Rock, California, tried the voucher plan, implementing it in the public school system with disappointing results and subsequently abandoning it.[10]

In 1971, the Panel on Non-Public Education of the Nixon administration's Presidential Commission on School Finance proposed "Parochiaid," which would have provided public money to religious schools. In the same year, the Supreme Court raised the legal barriers to government support for church schools. It held 8–0 in *Lemon v. Kurtzman* that distribution of tax dollars to private schools had to meet all of the following three tests to be constitutional: its purpose is secular; its main effect is to neither advance nor inhibit religion; and it does not excessively entangle the state with religion.[11,12]

Although "Parochiaid" died for lack of sufficient political support and the threat that it would be ruled unconstitutional, the idea of spending tax dollars on education at church–affiliated private schools remained alive. Indeed, the "Parochiaid" debate rehearsed many of the current arguments over private school vouchers and their use to pay tuition at religious schools.[13]

In 1983, 1985, and 1986, the Reagan administration tried unsuccessfully to move voucher legislation through Congress. By turning the federal government's means-tested Chapter 1 program into an individual voucher program,[14] the 1985 effort sought to re-establish the link between vouchers and "empowering" the poor, which had attracted liberals in the 1960s and 1970s.

## Educational Choice Enters The Mainstream

According to George Washington University Professor Jeffrey Henig, with free-market arguments for private school vouchers meeting with no success, the administration of President Reagan shifted the discussion to public school choice.[15] This new emphasis broadened support for school "choice," which many now saw as a strategy to reform rather than to dismantle the public school system. Furthermore, supporters often associated choice with educational excellence and racial equity through its link to the popular magnet school concept. Many school districts had established magnet schools to promote school integration and as an alternative to court-ordered busing. Magnet schools offered a diverse array of innovative curricula to attract voluntary transfers to integrated schools. By shifting the focus from private school vouchers to public school choice, President Reagan successfully separated educational choice from its racist and sectarian roots.[16]

Over the next eight years, beginning with Minnesota in 1988, 14 states enacted public school choice laws.[17] These laws allowed students to choose to attend any public school in the state that had room for them.

The idea of private school vouchers took the national stage again during the presidency of George Bush. Between 1990 and 1992, President Bush sent Vice President Dan Quayle to Oregon to speak on behalf of a voucher ballot initiative there. Bush expressed strong (and well-publicized) support for Wisconsin's 1990 private school voucher law, included "parental choice" in his 1991 "America 2000" reform initiative, and, in 1992, proposed a voucher plan he called a "G.I. Bill for Children."[18] Bush's Democratic challenger, Bill Clinton, took over the Reagan administration's "public school choice" position during the 1992 presidential campaign.

At the state level, private school vouchers have been vigorously debated for 20 years. Since 1978, four states have held referenda on voucher plans: Michigan (1978), Oregon (1990), Colorado (1992) and California (1993). Each of these efforts failed by an approximately 2 to 1 margin. California voters also rejected "regulated" voucher plans in 1980 and 1982 ballot initiatives.[19]

In 1993, Puerto Rico passed legislation that provided vouchers worth $1,500 per child that low-income families could use to send their children to any school, public or private (including religious schools that would accept them). The Puerto Rico Supreme Court struck

down the private school portion of the bill in 1994.

In 1995 and 1996, voucher legislation was introduced in Arizona, California, Colorado, Connecticut, Delaware, Florida, Illinois, Indiana, Minnesota, North Carolina, Ohio, Oregon, Pennsylvania, and Vermont. In addition, constitutional amendments were proposed in Michigan and Missouri to permit the creation of voucher plans.

At the federal level, a number of voucher proposals were recently introduced in Congress, including, in 1997, S.1 (the Safe and Affordable Schools Act), HR 1031 (the American Community Renewal Act), HR 2746 (the HELP Low-income Parents Act), and HR 1797 (the District of Columbia Student Opportunity Scholarships Act; the Senate companion bill is S. 847). The Washington D.C. appropriations bill for fiscal 1998 contained $7 million to establish a voucher experiment in the nation's capitol. As part of an agreement that led to the removal of voucher language from the D.C. appropriations bill, the Senate voice-voted its approval of a new voucher bill, S1502. S1502 would also appropriate $7 million for a voucher experiment. The House may vote on this bill as early as February or March 1998.

## The Battle Over Vouchers Today

Proponents of vouchers today base their position on three widely held views about public education: that educational outcomes have deteriorated, that American public education costs have accelerated unreasonably, and that the public schools cannot reform themselves because of bureaucratic and political constraints.

Notwithstanding the conventional wisdom, educational outcomes have actually improved. Between the 1970s and 1990, according to a 1994 RAND study, reading and math scores rose significantly for Hispanics and African-Americans.[20]

The best available evidence also shows that resources for regular classrooms at public schools have increased only modestly. In a survey of nine school districts, Richard Rothstein found that real spending for regular education climbed by only 28 percent from 1967 to 1991.[21] In Los Angeles, real per-pupil spending on regular education declined 3.5 percent over the same period. As Rothstein points out, if this decline typifies developments in urban areas generally, that may help explain frustration with academic outcomes.

Of course, national statistics about gradually improving performance, and the stagnation of funds to urban school districts, are of little comfort to parents convinced that their own children will not get the lift they need from the local public school. Parents who want better schools for their kids now have been a receptive audience for the third widely held view behind support for vouchers today: that public schools are incapable of reforming themselves because of bureaucratic and political constraints. This argument gained intellectual legitimacy with the publication of *Politics, Markets, and America's Schools* by John Chubb and Terry Moe, in 1990.[22] In their book, Chubb and Moe argue that the failure to improve school performance, despite a series of reforms instituted after the publication of *A Nation at Risk*, plus evidence of the superior performance of private schools, demonstrate the need for vouchers.[23] (For a summary of the public vs. private school literature, see Box 3).

The steep decline in the wages of male minority workers since the late 1970s has increased the urgency of demands to improve urban school quality and made many African-Americans receptive to vouchers. In Pennsylvania since 1979, with manufacturing jobs declining and non-professional employment stagnating in high-wage "bureaucratic" service industries (e.g., utilities, the telephone industry, the public sector), the median wage of African-American male workers plummeted by $3.59—from $12.72 in 1979 to $9.13 in 1996 in inflation-adjusted dollars.[24]

Many proponents of private school vouchers, such as Democratic Wisconsin Assembly member Annette "Polly" Williams, author of the Milwaukee Parental Choice Program legislation, have linked vouchers to their desire to empower poor families and raise the academic achievement of poor children. They argue that vouchers will improve achievement levels by forcing the public schools to compete in an educational marketplace in which poor parents hold the power of the purse.

KEYSTONE RESEARCH CENTER

# The Milwaukee Parental Choice Voucher Program

Until the Wisconsin state legislature passed Act 36 in 1990, establishing the nation's first private school voucher program, the debate over vouchers took place wholly on the ideological and philosophical plane. Even today, the Milwaukee Parental Choice Program (MPCP) is the only voucher program for which large amounts of systematic data are available. For this reason, the Milwaukee program occupies a central place in any discussion of the merits of private school vouchers.

The MPCP initially allowed up to 1 percent (about 1,000) of low-income Milwaukee Public School students to attend participating private, non-sectarian schools within the city. The program defined "low-income" as below 175% of the official U.S. poverty line. Each child attending a private school in the program receives a voucher worth the per-pupil equalized state aide to the Milwaukee Public Schools, originally set at $2,446.

Participating schools had to meet only one of four educational requirements:

1) at least 70 percent of pupils advance one grade level each year,

2) attendance averages at least 90 percent,

3) at least 80 percent of students demonstrate significant academic progress, or

4) at least 70 percent of their families had to meet parental involvement criteria established by the private school.

Unlike public schools, teachers at Choice schools need not be certified, nor does the curriculum of the schools have to be reviewed or accredited by an outside agency. Choice schools do not have to meet the financial disclosure or other record keeping requirements placed on the public schools. After a lawsuit, participating private schools need not serve children with exceptional educational needs.

The Wisconsin legislature created Milwaukee's Choice program as a five-year experiment and provided for yearly evaluations of the academic achievement of students attending Choice schools. Governor Thompson vetoed the five-year time limit on the program but left the requirement of annual program evaluations intact. The Wisconsin Supreme Court upheld the constitutionality of the Wisconsin law in 1992 reasoning that it affected a small number of children living in poverty, did not include religious schools, and what the state learned from the experience might benefit children elsewhere in Wisconsin.[25]

In 1993, Act 16 modified the Milwaukee Parental Choice Program to raise (effective 1994–95) the number of students who could participate from 1 percent to 1.5 percent (about 1,500 students) of the Milwaukee Public School (MPS) population. The same Act allowed the maximum number of Choice students at participating schools to increase from 49 percent to 65 percent of the total student population.

Since 1990, there have been five official yearly evaluations of the Milwaukee voucher experiment (discussed at length in the next section) by University of Wisconsin political science Professor John Witte.[26] Witte found no statistically significant differences between the achievement of students attending Choice schools and the achievement of random samples of students attending the Milwaukee Public Schools. He did, however, find a high degree of parental satisfaction with Choice schools.

A 1995 report by Harvard Professor Paul Peterson sharply criticized Witte and his statistical methods.[27] These methods, Peterson argued, understated the positive academic impact of the Milwaukee Parental Choice Program. Peterson's argument echoed a 1992 critique, "The Milwaukee Parental Choice Program," written by George Mitchell for the Wisconsin Policy Research Institute.[28]

In February 1995, the Wisconsin Legislative Audit Bureau, the research arm of the legislature, released its own report on the Milwaukee

program. The report did not find Witte's methods inappropriate. However, it contended that no conclusion—not even Witte's finding of no significant difference—could be drawn about academic performance under the voucher program compared to the Milwaukee Public Schools.[29]

During the 1995 legislative debate over the expansion of the Choice program, the Peterson critique and Witte annual reviews enabled both advocates and opponents to claim that the data supported their position. Unfortunately, instead of attempting to strengthen and improve the evaluation requirements for the Milwaukee Parental Choice Program, voucher supporters lobbied successfully to eliminate the annual program evaluation requirement. As revised in 1995 (Act 27), the evaluation components of the MPCP consisted of a requirement that the Legislative Audit Bureau report on the finances and performance of the program after five years (January 15, 2001) and a provision requiring that each voucher school provide the Wisconsin Department of Public Instruction with an annual independent financial audit. The 1995 revision of the MPCP did not, however, require that the schools participating in the program gather the achievement data necessary for a rigorous evaluation.

The 1995 legislation allowed religious schools to participate in the program, and raised the number of students who could participate to 7 percent of the Milwaukee Public School enrollment in 1995–96 and 15 percent in 1996–97. The new legislation also allowed up to 100 percent of the students attending a Choice school to be voucher students.

On August 25, 1995, the Wisconsin Supreme Court enjoined all of the 1995 modifications to the Milwaukee Parental Choice Program. On March 29, 1996, the supreme court deadlocked 3–3 on the constitutionality of 1995 modifications and sent the case back to circuit court for trial. On August 15, 1996, the circuit court retained the injunction barring implementation of religious school participation in the program but lifted the injunction on other parts of the 1995 legislation. The Dane County Circuit Court ruled the entire 1995 Act unconstitutional on January 15, 1997. An appeal is currently before the Wisconsin Supreme Court. As of the 1997–98 school year, the 1993 modification to the 1990 law again governs the MPCP.

As a result of the changes enacted in 1995 and subsequent court actions, no achievement data on the MPCP were collected during the 1995–96 or 1996–97 school years. During 1997–98, the evaluation requirements built into the original law govern the program. This may change when the Wisconsin Supreme Court issues its ruling in the spring of 1998.

KEYSTONE RESEARCH CENTER

# The Debate Over The Achievement Effect Of The Milwaukee Voucher Program

Three research teams have analyzed the data collected during the first four years of the Milwaukee voucher program.

- University of Wisconsin-Madison political science professor John Witte is the principal author of each of the first four annual evaluations of the program.[30] He and his team are the only researchers to have analyzed fifth-year data on the program.[31] In a January 1997 paper, Witte summarized the findings of his first four evaluations and presented a reanalysis of some of his data in light of criticisms of his methods and findings.[32]

- In August 1996 and March 1997, Professors Jay Greene (University of Houston), Paul Peterson (Harvard) and Jiangtao Du (Harvard) issued two reanalyses of Witte's data on the first four years of the program.[33]

- In September 1997, Princeton Professor Cecilia Rouse released a paper, accepted for publication in the *Quarterly Journal of Economics*, that analyzes the achievement data from the Choice program's first four years.[34] In December 1997, Rouse published a subsequent paper comparing performance in three categories of schools within the MPS system, both to each other and to the Choice schools.

In considering the research designs and findings of Witte, Greene, Peterson, and Du, and Rouse it is useful to understand the Milwaukee Parental Choice Program's scope and character. The program has never involved a large number of students and has never reached the total enrollment authorized by law. Some students have nonetheless been turned away because the school they wished to attend had no space at their grade level. According to the Wisconsin Legislative Audit Bureau's 1995 report, 30.3 percent of the children enrolled in the program one year do not return the next year.[35]

**Table 1**
**Milwaukee Parental Choice Program Profile**

| | Number of Schools | Number of Applications | Average # of Voucher Students* | Voucher Amount | Total Cost of Vouchers (millions) | Annual Attrition Rate |
|---|---|---|---|---|---|---|
| 1990-91 | 7 | 577 | 300 | $2,446 | $0.73 | 0.46 |
| 1991-92 | 6 | 689 | 512 | $2,643 | $1.35 | 0.35 |
| 1992-93 | 11 | 998 | 594 | $2,745 | $1.63 | 0.31 |
| 1993-94 | 12 | 1049 | 704 | $2,985 | $2.10 | 0.27 |
| 1994-95 | 12 | 1046 | 771 | $3,209 | $2.47 | 0.28 |
| 1995-96 | 17 | | 1288 | $3,667 | $4.60 | |
| 1996-97 | 20 | | 1616** | $4,373 | $7.07** | |
| 1997-98 | 23 | | | $4,696 | | |

*Includes summer school.
**Unaudited figures.
Sources: State of Wisconsin Department of Public Instruction web page, http://www.dpi.state.wi.us/dpi/dfm/sfms/histmem.html; and John F. Witte, Troy D. Sterr, and Christopher A. Thorn, *Fifth-Year Report: Milwaukee Parental Choice Program* (Madison, WI: and The Robert M. La Follette Institute of Public Affairs, University of Wisconsin-Madison, December 1995).

The MPCP overwhelmingly supports elementary school students. According to the 1995 Legislative Audit Bureau Report, 23.2 percent of the participants in the Milwaukee voucher program in 1994–95 enrolled in kindergarten, 61.1 percent in kindergarten through third grade, and 76 percent in kindergarten through fifth grade.[36]

For 1997–98, a MPCP voucher equals $4,696.[37] The Milwaukee Public Schools also provide transportation for those voucher students who require it. The voucher compares with a per-pupil expenditure in the Milwaukee Public Schools of $7,869 for 1997–98. (As well as the state support that sets the voucher amount, MPS total spending per pupil includes funding from local tax revenues, federal aid, and private sources.) Of the $7,869 total, on average, elementary (K-6) schools directly received $3,875 per-pupil, K-8 schools received $4,234, middle schools $4,831, and high schools $4,659 per pupil. Over and above these amounts, schools also receive money for special education. Money not distributed directly to the schools is used for capital improvements, the recreation program, alternative education programs, food service, building maintenance, transportation, and other central support services. Central administration costs account for approximately 5 percent or less of the Milwaukee budget.[38]

In sum, while Brent Staples in *The New York Times* claimed on January 4, 1998, that vouchers are limited to $3,000 and are less than half what public schools spend per pupil, neither statement is true.[39] Indeed, since Choice students fall primarily in the relatively inexpensive primary grades, vouchers usually exceed what most MPS schools receive directly for pupils in the same grades. It is impossible to judge whether voucher or public schools have more resources in Milwaukee at this juncture, because information is lacking on what participating private schools receive from private sources, and because the range of services offered by private and public schools differs (private schools, for example, need not provide special education).

Three schools, Bruce Guadalupe, Harambee, and Urban Day, enroll a substantial majority (over 80 percent according to Greene, Peterson, and Du [40]) of all voucher students. Each of these schools had a long history and established reputation prior to the passage of the Milwaukee voucher program. The fact that three schools, with unique histories, enroll such a large proportion of Milwaukee's voucher students makes it difficult to generalize to large-scale voucher programs that would require many new schools. Finally, none of the evaluations of the Milwaukee program contain data on high school students because so few voucher students attend high school.

In his evaluations, John Witte found that, when compared to Milwaukee Public School parents, parents who send their children to voucher schools are better educated and more involved in their children's education, have higher academic expectations, and are more critical of the Milwaukee Public Schools than are Milwaukee Public School parents.[41] These findings have not been disputed. This suggests that MPCP parents are so-called high-voice parents. Since only a small number of students apply to Choice schools each year (see Table 1) relative to the number of eligible students (about 60,000), the program may be attracting a small subset of low-income parents with distinct characteristics. This makes it difficult to use the Milwaukee experience to predict the effectiveness of large-scale voucher programs.

To determine the academic impact of the Milwaukee voucher program, all of the researchers whose work is described here use test data from the Iowa Test of Basic Skills in reading and math.

KEYSTONE RESEARCH CENTER

## BOX 3: PUBLIC VS. PRIVATE SCHOOLS

Both Rouse and Greene, Peterson, and Du locate their Milwaukee voucher program research within the literature on private vs. public school performance. Much of this literature begins from the premise that private schools are better at responding to competition than public schools and are therefore likely to be more efficient at producing desirable educational outcomes.

Studies both support and refute the premise that private schools are better at producing high achieving students. Evans and Schwab,[42] for example, found overall positive effects from attending Catholic schools, while Goldhaber found no advantage of private school attendance.[43] One of the most contentious issues in this research literature is the issue of selection bias, i.e., whether differences in achievement are explained on the basis of who attends private schools. The unrepresentative set of private schools in one widely used data base (High School and Beyond) is also of concern.

In a recent study, David Figlio and Joe Stone of the University of Oregon drew on the National Education Longitudinal Survey and a Dun and Bradstreet directory of private schools to analyze public and private school performance in 8th-12th grade math and science.[44] Their research attempts to simulate the placement of otherwise equivalent students into different school environments, and thereby to isolate the achievement effect of attendance at a public vs. private school. Figlio and Stone caution that their results on the performance of low-income and low-achievement students are based on very small numbers (47 low-income students and 39 low-achieving students).

Figlio and Stone's study reveals the complexity of the issue of private vs. public school performance and the danger of drawing simplistic, sweeping conclusions about the relative performance of public and private schools. Figlio and Stone estimate either no achievement effect or negative effects overall for attendance at a religious school. They find, however, that African-American and Hispanic students who attend religious schools outperform their public school counterparts, especially in urban areas. According to Figlio and Stone, non-religious private schools have a positive effect on math and science achievement primarily for low-income and initially low-achieving students. High-achieving students may do less well in science in private non-religious schools.

Figlio and Stone advise that their findings should be used very carefully if deployed in the debate about vouchers. As they explain, their estimated effects only simulate what would happen if a few students moved from private to public school. In this situation, when low-income and initially low-achieving students attend private schools, these students may benefit from changes in who is in school with them—"peer group composition." What Figlio and Stone cannot estimate is the effect on achievement that would occur if larger numbers of students moved from public to private schools. This would cause large changes in peer group relationships at both sending and receiving schools. Large-scale implementation of vouchers could have negative achievement effects in both public and private schools because of the changes in student body composition it could produce.

On the whole, the research literature gives no clear guidance as to whether or not private schools are better at producing desired educational outcomes than public schools. Since most of the studies use data for secondary schools, they are of limited value in understanding the impact of voucher programs that involve elementary schools.

KEYSTONE RESEARCH CENTER

## Why Different Researchers Reach Different Conclusions

When researchers in ideologically polarized debates disagree, general readers who want to weigh the "facts" for themselves can end up confused and not knowing what to think. To avoid this problem, this section walks the reader through the findings of the three efforts to analyze the Milwaukee experience. It seeks to explain in everyday language how essentially the same underlying data can lead different analysts to different conclusions.[45]

There is actually less disagreement than meets the eye between the findings of the three Milwaukee evaluations. When researchers of the MPCP program use similar methods, they come to the same basic conclusions.

Researchers of the Milwaukee voucher program arrive at conflicting results for two basic reasons: (1) they use different definitions of the reference or control group to which the performance of voucher program participants should be compared, and (2) they use different methods to control for family background and student ability. All of the researchers must contend with the relatively small samples of students in the data bases analyzed. All must address the shrinkage (or "attrition") of their sample due to student mobility and missing data. *All of them also lack any model of what actually goes on in schools or of the educational features (such as small class size or an innovative curriculum) that may generate good outcomes.*

**Table 2**
**Findings of Three Studies of the Milwaukee Parental Choice Program**

| | Witte | Greene, Peterson, and Du | Rouse |
|---|---|---|---|
| **Main Comparison**[46] | Compares voucher students' achievement with that of a random sample of Milwaukee Public School (MPS) students, controlling for observed individual and family characteristics. | Compares voucher students' achievement with that of unsuccessful applicants who returned to the Milwaukee Public Schools. | Compares achievement of successful applicants for vouchers with that of a random sample of Milwaukee Public School students, controlling for an estimate of innate ability and family influences. |
| **Reading Findings** | No significant difference between voucher students' achievement and that of the MPS comparison group. | In their 1997 "main analysis": 2–3 percentile rank advantage for voucher students' in year four. Conventional levels of statistical significance approached only when 3rd and 4th years are jointly estimated. When background characteristics are controlled for, voucher students' advantage in 1st and 3rd years approaches significance. | Similar to Witte: no statistically significant difference between successful voucher applicants' achievement and that of the MPS comparison group. |
| **Math Findings** | No significant difference between Choice students and MPS sample. | 5–11 percentile rank advantage for voucher students over unsuccessful choice applicants in years 3 and 4. Conventional levels of statistical significance achieved in 4th year and in joint estimate of 3rd and 4th years. | Similar to GPD: statistically significant advantage in years 3 and 4 for students selected for Choice schools. Effect size of 0.08–0.12 per year. |
| **Main Statistical Limitations** | • Does not control for unobserved individual differences.<br>• Voucher students who remain in program may be a non-random high-scoring group.<br>• Does not include school variables (e.g., class size, curricula). | • Control group of unsuccessful voucher applicants who return to MPS is a small and shrinking sample (26 in year 4).<br>• Control group may be a non-random, low-scoring group.<br>• Voucher students who remain in program may be a non-random high-scoring group.<br>• Does not include school variables (e.g., class size, curricula) that may explain observed differences. | • Successful voucher applicants have more educated parents with high expectations: improvement in math scores over time might take place without voucher program.<br>• Does not include school variables (e.g., class size, curricula) that may explain observed differences (see text and Box 11). |

KEYSTONE RESEARCH CENTER

## The Witte Evaluations

In his five evaluations of the Milwaukee program, Witte compares voucher students' average test scores and changes in test scores to the same figures for two other groups: a random sample of Milwaukee Public School students and a random sample of low-income Milwaukee Public School students. Since neither of these two groups are genuine "control" groups for Choice students, Witte also combines the Choice and non-Choice students into a single sample and uses statistical controls to take account of the impact of family and individual differences (e.g., prior test performance, family income, race, and gender) on test scores.

### BOX 4: SORTING THROUGH CONFLICTING VOUCHER RESULTS

To help you avoid getting lost in the technical summary of voucher research on Milwaukee, the list below summarizes this report's distillation of what the research tells us.

- Disagreement exists about whether the voucher program generates outcomes compared to the Milwaukee Public School (MPS) system. Two of three research teams think no positive outcomes result in reading. Two of three teams think that positive outcomes result in math.

- The evaluations all deal with small samples. Many students drop out of the experiment, possibly on a non-random basis. These data deficiencies should be kept in mind when interpreting the results.[47]

- The parents of voucher applicants have more education and higher expectations than parents of most Milwaukee Public School students. Wherever they attend school, the children of such parents may improve over time compared to other students.

- Students in a group of public schools with small classes outperform Choice students (according to the only analysis that looks at different groups within the MPS system).

- Lacking the necessary data, the evaluations cannot look at the educational process inside the Choice schools. They cannot explain what lies behind any differences in performance between Choice and MPS schools or among the Choice schools.

- Over 80 percent of Milwaukee voucher students attended three schools with established reputations. At best, the experiment tells us something about how these particular private schools compare with Milwaukee public schools, as a group. It indicates nothing about the impact of larger-scale voucher programs.

Taking account of these differences requires including in the analysis only students for whom there are complete data, which exacerbates the problem of sample size.

Witte's overall conclusion: there is no academic advantage for students attending Choice schools. He finds a small, non-significant advantage for Milwaukee Public Schools in reading.[48]

## The Greene, Peterson, and Du Evaluation

Greene, Peterson, and Du (GPD) argue that, when Witte compares Choice and MPS students, his controls for family and individual characteristics are inadequate.[49] Therefore, GPD choose a method different from Witte's.[50] They compare Choice students to students who applied to but did not get into Choice schools. The Milwaukee voucher law required that each participating school randomly select its successful voucher applicants. GPD therefore consider a comparison of successful and unsuccessful applicants to be akin to a natural experiment comparing two otherwise identical groups. In their view, differences that may exist between students do not have to be controlled for because random assignment assures that differences will be evenly distributed across the groups being compared.

Several factors mar their natural experiment, however:

- First, no one has examined whether Choice schools actually selected randomly. (In response to this point, GPD show the prior test scores and family characteristics of the two groups to be similar "in essential respects."[51])
- Second, siblings of children already enrolled in Choice schools were guaranteed places without going through the lottery.
- Third, since lotteries took place at the school level, each school's group of Choice students has its own control group of rejected applicants. The available data, however, does not indicate the particular Choice school to which unsuccessful applicants sought admission. To model the lottery process, GPD therefore assume that Hispanic students applied to the predominantly Hispanic school and that African-Americans applied to one of the two other schools with large numbers of voucher recipients. This technique leads GPD to leave white students out of the analysis.

Aside from questions about the randomness of the original selection process and the difficulties of modeling it, a number of other problems result from GPD's reliance on unsuccessful Choice applicants as a comparison group. First, only a relatively small number of applicants failed to get into the voucher program each year (see Table 1). Moreover, many of these applicants dropped out of the Milwaukee Public Schools by the third or fourth year of the program, aggravating GPD's sample size problems. The largest number of Choice students analyzed by GPD in the third year is 310, with only 86 in the control group. By the fourth year, the largest number of Choice students analyzed by GPD is 110, with only 26 in the control group. This makes the estimated effects unusually sensitive to a few very high or low scores.

As Witte and Rouse note, moreover, unsuccessful Choice applicants who returned to the Milwaukee Public Schools are not only a smaller group over time, they may also be progressively less representative. In part because of the availability of a privately funded voucher program (see the discussion of PAVE below), many unsuccessful applicants found the resources to leave MPS. Those remaining in MPS may constitute an atypical, low-performance sub-group, particularly in years three and four. Consistent with this possibility, after four years, the family income of unsuccessful Choice applicants remaining in the MPS system is over $6,500

below that of unsuccessful applicants who leave MPS. The parental education of those still in MPS also falls slightly below that of the group who left.[52]

While unsuccessful applicants may be a low-performance group, the opposite may be true of those left in Choice schools in later years. (This problem plagues Witte's analysis as well as GPD's.) GPD themselves report evidence that voucher students who remain in the program are an unrepresentative, high–performance group (see the last part of Box 5). University of California-Berkeley Professor Bruce Fuller suggests that drawing a conclusion from looking at students left in Choice schools would be like determining the effects of smoking by only tracking smokers who didn't die.[53]

Comparing Choice students to unsuccessful Choice applicants, GPD report that, after three or four years in the Choice program, students begin to show higher levels of performance. In math, GPD report 5 and 11 percentile rank differences in the third and fourth years.[54] Reading scores of Choice students exceed those of unsuccessful applicants by 2 to 5 percentile ranks. GPD say that the delay before math and reading scores improve may result from the time it takes students to accustom themselves to a new school and its academic program.

When GPD take account of students' individual characteristics on which they have data, their results achieve conventional levels of statistical significance once and approach significance six other times. GPD maintain, as noted earlier, that random selection of voucher recipients from each school's applicant pool means that there is no need to control for individual characteristics. While random assignment does mean that individual characteristics should not make much difference, it does not justify excluding them. GPD counter that the lack of statistical significance of their results (once they include background characteristics) results not from any reduction in the positive impact of Choice schools, but rather from a reduction in the sample size because the data do not contain complete information on individual characteristics for all students.

In 1997, following GPD's analysis, Witte himself looked at the performance of unsuccessful Choice applicants.[55] In reading, he finds, Choice students perform no differently than unsuccessful applicants. In math, like GPD, Witte finds that Choice students do better than unsuccessful applicants, especially in the third and fourth years in the program. Witte, however, discounts the value of these results because 52 percent of unsuccessful applicants did not return to MPS, so no test scores are available for them. He argues that the remaining unsuccessful applicants do not constitute a random sample of unsuccessful applicants. Witte also suspects his math results because his total sample for this comparison includes only 85 students who had been in the Choice program four years, and only 27 unsuccessful applicants. Moreover, the achievement difference can be accounted for by the scores of only five unsuccessful applicants who did not appear to answer any of the test questions. When Witte eliminates the scores of the lowest scoring group of students (five unsuccessful applicants and two Choice students), he finds that the math effect was no longer statistically significant. Moreover, the unsuccessful applicants did even more poorly against a random group of MPS students than against Choice students.

Based on their results, Greene, Peterson, and Du speculate that vouchers, if generalized and extrapolated to all white and minority students in the United States, would eliminate most of the achievement gap between white and minority students in reading and erase it altogether in math. It is not clear on what grounds GPD base this speculation because they exclude all white students from their analyses.

Greene, Peterson, and Du's overall conclusion: participation in the Milwaukee Parental Choice Program confers academic achievement advantages in reading and in math that are cumulative and that first appear after three years in the program.

KEYSTONE RESEARCH CENTER

## BOX 5: WHEN ARE SIGNIFICANT RESULTS NOT SO SIGNIFICANT?

In statistical analysis, social scientists need to know how to distinguish findings that could be the result of random chance from findings that indicate strong confirmation of a hypothesis—such as the hypothesis that Choice schools improve student performance. By convention, social scientists most commonly consider a result "statistically significant" when the probability of it occurring by chance is .05 (i.e. 5 chances out of a 100) or less. In their March 1997 paper, however, Greene, Peterson, and Du report a result as significant when there is a 1 in 10 (or .10) probability or less of it occurring by chance.

GPD further increase the number of "significant" findings that they report by evaluating results using a "one-tailed" test of significance rather than a more common "two-tailed" test.

One-tailed tests are usually used when there are strong theoretical reasons for believing that change in the independent variable (in this case attendance at Choice schools) is likely to produce a change in the dependent variable (test scores) in only one direction. GPD's theory is that Choice students could not perform worse on tests than those who applied to the program but were rejected. GPD justify this by reference to the literature suggesting that private schools perform better than public schools. It is a questionable assumption because, as we saw in Box 3, the literature on private vs. public school achievement is drawn primarily from secondary school data, shows mixed results, and is very controversial. Rouse's finding that students in a sub-group of Milwaukee public schools outperform those in Choice schools raises further questions about the one-tailed assumption.

The important point here is that by using both a .10 standard of significance and a one-tailed test in their March 1997 paper, GPD are four times more likely to find significant results than if they had applied a .05 standard using a two-tailed test. This allows them to report almost eight times as many statistically significant finding in Tables 3, 4, 5, and 7 of their March paper than they would have been able to report using a .05 level with a two-tailed test. In other words, they report 23 significant findings instead of 3.

GPD might respond that "statistical significance is not a cliff" and that results slightly below a customary threshold for significance are still unlikely to occur by chance and are therefore worthy of note. GPD, however, are not consistent in this view. In one important case, they fail to point out some significant findings (at the .10 level) that reduce confidence in their main finding about the performance advantage of voucher students. This case comes up when GPD respond to the claim that lower-performing students more often leave the voucher program, making their sample of students still in the Choice schools unrepresentative. In their August 29, 1996 paper, GPD directly test for such attrition bias by comparing (a) the scores of students who continued in the voucher program with (b) the scores of students who withdrew from the program (i.e., the last score of these students before they left the voucher program). GPD summarize their findings as follows:

> In only two comparisons were differences statistically significant. In one the students leaving the study had the higher test scores; in the other, continuing students had higher test scores. In the other six cases, the two groups did not differ significantly.

When you look in their table reporting these results (Table 7 in their paper), you find that two of the "insignificant" differences between Choice stayers and leavers are nearly significant (they could have occurred by chance with only a .06 and a .09 probability). These differences meet the .10 standard that GPD earlier used as a threshold for significance. In both these cases, the math scores of continuing choice students exceed the math scores of those who drop out of the program. Perhaps adding to the inconsistency, GPD may have used a two-tailed test in their examination of Choice student attrition bias. If one accepts the theory that more successful students in Choice schools would not leave the voucher program, then a one-tailed test would be more appropriate. Under a one-tailed test, the math advantage of continuing Choice students over those who quit in 1993 and 1994 would be significant at a .05 level.

KEYSTONE RESEARCH CENTER

## The Rouse Evaluation

The most recent analysis of the Milwaukee Parental Choice Program data has been done by Professor Cecilia Rouse of Princeton. Rouse analyzes the performance of all students selected to attend Choice schools (including those who never attended—a small group—and those who subsequently left). She compares this group's performance to that of applicants not admitted to the Choice program and to a random sample of MPS students. By comparison with GPD's main method, this approach has the advantage of avoiding non-random attrition from the Choice sample. It also increases the number of students in the "Choice" sample. Rouse sees including all those awarded vouchers in the Choice group as a better way of assessing the overall impact of the MPCP program than restricting the sample to those currently receiving vouchers. According to GPD, who use the same method in part of their March 1997 paper, Rouse's approach better captures what would happen if the Choice experiment were generalized and students migrated back and forth between private and public schools.

In addition to her analysis of successful applicants for vouchers, Rouse does a more familiar comparison between students who actually attended Choice schools and her MPS sample. Whichever way she defines program participants—as those selected or those actually attending—Rouse's estimate of their test scores relative to those of Milwaukee Public School students turns out to be similar.

Like Witte, Rouse finds no significant advantage for the Choice groups in reading. She describes the Greene, Peterson, and Du results for reading as "fragile."[56] In math, Rouse finds that students admitted to the voucher program, and the sub-sample still participating in it, both had faster math gains than her random sample of MPS students. She estimates that the math scores of successful applicants and of program participants rise each year by 1.5–2.4 percentile points more than MPS student test scores. This amounts to an effect size of 0.32–0.48 over four years (see box 2 for a definition of effect size).

Rouse argues that the difference between her and Witte's comparison of the math scores of MPS and voucher students results from a highly technical difference in the statistical models used. (She supports this claim by making her model similar to Witte's and showing that she gets results comparable to his.) While Witte's model includes prior test scores (and other individual characteristics) as controls, Rouse uses an individual "fixed effects" model that controls for all student characteristics that do not change over time (e.g., parental education and "innate" ability).[57] Rouse's approach enables her to include in her sample individuals that Witte excludes because of missing some prior year test scores.

Rouse cautions that there are several caveats to bear in mind when considering her results.[58]

- First, a large number of students in the data set do not have total math scores. (This is a problem for all three research teams.) For 1993, Rouse had to impute the total math score (from scores on the components of the test) for 40 percent of the unsuccessful Choice applicants and 34 percent of the students in her Milwaukee Public Schools sample. For 1994, she had to impute 69 percent of the total math scores for the unsuccessful Choice applicants and 67 percent of the Milwaukee Public School sample.

- Second, Rouse's method assumes that, in the absence of the voucher program, the two comparison groups would have improved their scores over time at the same rate. If, however, the test scores of children with high-voice parents tend to improve faster than the test scores of other students—even when the high-voice offspring start off poorly—then Rouse's model would wrongly attribute this improvement to the voucher program.

- Third, the data sets on the Milwaukee voucher experiment include no school variables, such as social and economic profile of the school, class size, school size, or spending

per student. Therefore, neither Rouse nor the other analysts have any way of knowing whether differences between the achievement of Choice students and that of Milwaukee Public School students are attributable to these variables. Since there is clear evidence that class size, for example, has a significant effect on student achievement, Rouse's results may have nothing to do with participation in the Choice program per se. In her most recent paper, analyzed at length in the class-size section of this report, Rouse takes a first step towards addressing the lack of school variables. She presents evidence that class size in public schools exceeds that in Choice schools. Moreover, she finds that students in the one sub-group of the Milwaukee Public Schools that have a class size comparable to Choice schools have better overall test scores than Choice schools.

- Finally, Rouse points out that the average effects she reports say nothing about the performance of individual Choice schools, i.e., they do not suggest that all Choice schools are "better" than the Milwaukee Public Schools.

Rouse's overall conclusion: allowing low-income children to attend private schools might raise the math achievement of those who participate. However, the Milwaukee data do not answer the question of whether vouchers give public schools an incentive to improve, nor do these data provide an adequate basis for making decisions about the widespread implementation of voucher programs.

Rouse ends her December 1997 paper by noting:

> *If we really want to "fix" our educational system, then we need a better understanding of what makes a school successful, and not simply assume that market forces explain sectoral differences and are therefore the magic solution for public education.*[59]

KEYSTONE RESEARCH CENTER

## Milwaukee's Private Voucher Program—PAVE

Voucher programs supported by private sources provide another potential source of information on the educational consequences of vouchers. Perhaps the country's largest private program operates in Milwaukee. Partners Advancing Values in Education (PAVE), formerly the Milwaukee Archdiocesan Education Foundation, was founded in 1992. PAVE provides low-income families with scholarships worth half of the tuition charged by a private religious or non-sectarian school up to a maximum of $1,000 for elementary and middle school students and $1,500 for high school students. PAVE's overhead is about 7 percent of its annual costs.[60]

PAVE awards about half of its scholarships to students who already attended private school. Approximately 95 percent of PAVE-supported students attend religious schools, with more than half (about 60 percent) enrolled in Catholic schools. Unlike the Milwaukee Parental Choice Program, PAVE enrolls a higher percentage of white students than the Milwaukee Public Schools. Also, unlike MPCP, schools participating in PAVE may reject applicants.[65]

PAVE has for the most part shied away from assessing student achievement gains preferring to focus on other issues such as parental satisfaction, parents' reasons for participating in PAVE, and the extent to which they assist with their children's school activities.[66] The most recent (1996) evaluation, for example, examined discipline in participating schools, the residential mobility of participating families, and the reasons eligible families did not participate.[67] The evaluations commissioned by PAVE have found that people who participate in the program are well satisfied and that there are relatively few serious discipline problems at PAVE schools.

Of the four evaluations of the PAVE program, only the 1994 report made a serious effort to determine the program's effect on student achievement. The 1994 evaluation suggested that students who attended private schools for their entire school career achieved at higher levels than students who transferred from a public school into a private school participating in the PAVE program. Further, the evaluation suggested that the longer transfer students stayed in participating private schools the greater their achievement.

Unfortunately, since the data gathered depended entirely on the voluntary cooperation of parents, the findings are suspect and no conclusion can be drawn from the evaluation's results.

### BOX 6: MILWAUKEE – A CASE EXAMPLE OF THE RELATIVE COST AND PERFORMANCE OF PUBLIC AND PRIVATE SCHOOLS

Milwaukee provides a case example on both the relative performance and the relative cost of public vs. private schools. In 1991, the Catholic archdiocese of Milwaukee released the test scores of children in its schools. The results showed that when the performance of children from similar social and economic backgrounds were compared, the Catholic schools in the Milwaukee archdiocese did no better and perhaps a bit worse at educating minority children than the Milwaukee Public Schools.[61]

The picture looks about the same with the issue of cost. In 1994, when the archdiocese began closing its four central-city elementary schools, the Catholic school system had a per-pupil cost of approximately $4,000 at the four schools.[62] By comparison, in the 1992-93 school year, when excluding centrally budgeted items such as fringe benefits and transportation, each elementary school in Milwaukee received, on average, $2,958. Even including all centrally budgeted items the public schools spent $4,645 per student.[63] The Milwaukee public schools also provide many more services and a more complete educational program than the private Catholic schools, according to an independent Milwaukee-based research institution.[64]

KEYSTONE RESEARCH CENTER

# The Cleveland Scholarship And Tutoring Program (CSTP)

Ohio enacted the Cleveland Scholarship and Tutoring Program (CSTP) legislation into law in March 1995.[68] It allowed the Ohio Superintendent of Public Instruction to create a pilot voucher program in Cleveland. It was expected that the $6.4 million appropriated for the program's first year would be enough for 1,500 scholarships. The Cleveland program is largely supported by $5.25 million from Ohio's Disadvantaged Pupil Impact Aid Program previously earmarked for the Cleveland Public Schools.

For families whose income is less than double the Federal poverty level, CSTP provides vouchers of up to 90 percent of a private school's (including religious schools) tuition, up to a maximum of $2,250. If a family's income is more than twice the Federal poverty level, the state pays up to 75 percent of a participating school's tuition to a maximum of $1,875. Up to 25 percent of the new scholarships each year may be awarded to children previously enrolled in a private school.

Scholarship applicants are selected by lottery with priority going to applicants whose income is less than the Federal poverty level. Second priority goes to families whose income is less than twice the poverty level. Within these guidelines there is no income cap on participation.

The approximately 30,000 K-3 students who reside within the Cleveland School District are eligible to apply to the program. Once admitted to the program, students may receive scholarships through 8th grade. In the first year, 6,246 applications were received for the 1,500 slots assigned by lottery in January 1996. Over the next several months, the state increased the number of vouchers that could be awarded to 1,801 because it more accurately calculated the actual tuition amounts involved. Ultimately, all public school applicants were offered a voucher. However, there was a waiting list of students previously enrolled in private schools. At the start of the 1997–98 school year, the total number of participants increased to 3,000.

In 1996–97, about 35 percent of the participants were kindergartners with no previous enrollment history, another 35 percent were formerly enrolled in the Cleveland Public Schools, and about 29 percent (up from 25 percent because of lower attrition among students already in a private school) were previously enrolled in private schools. Since some kindergarten students would have enrolled in private school even without the program, the 29 percent figure is probably a conservative estimate of the share of voucher recipients that would be in private schools anyway.

In 1996–97, about 77 percent of the scholarship students attended one of 46 religious schools, 35 of which are Catholic. The other 23 percent attended non-sectarian private schools, with over three quarters of them attending two schools. Although the law allows program participants to attend suburban public schools, none did. The vast majority of participants in the program are low-income African-Americans.

The actual cost of Cleveland scholarships to taxpayers is somewhat controversial. The Ohio Office of Management and Budget sets the average voucher payment for 1996–97 at $1,763. An analysis by the American Federation of Teachers estimates the cost of transportation at $629 per scholarship recipient, the cost of administering the program at $257 per student, and the additional state aid the program generates for each scholarship student enrolled in a private school at $543. Using these figures, the AFT estimates the total scholarship cost at $3,192 per recipient.[69]

The Cleveland program is a scholarship and *tutoring* program. By law, the number of Cleveland public school tutorial-grant recipients may not exceed the number of students who receive vouchers. The value of tutorial grants is based on an income-related sliding scale up to a maximum of 20 percent of the average scholarship amount (i.e., the tutorial grant ceiling equals $450 for families with income below twice the poverty line and so on). In 1996–97, 542 students received tutorial assistance and there was a waiting list of 201 students

who were unable to find a qualified tutor.

Since, unlike the Milwaukee Parental Choice Program, the Cleveland voucher program allows religious schools to participate, its constitutionality was immediately challenged. On July 31, 1996, the Franklin County Court of Common Pleas held the program constitutional and allowed it to be implemented. On May 1, 1997, an Ohio appeals court ruled the program unconstitutional. The Ohio Supreme Court allowed the program to go forward while it considers an appeal. Its ruling is expected early in 1998.

The Cleveland Scholarship and Tutoring Program legislation requires the state superintendent to contract with an independent research entity to conduct an evaluation of the program's impact on student performance, parental involvement, public schools, and the market supply of alternative education. The contract to evaluate the program was awarded to an Indiana University research team headed by Professor Kim Metcalf. An evaluation report on the program's first year is expected in early 1998.

There has been some confusion surrounding the Cleveland evaluation because of the publicity associated with the analysis of test score data from the two largest non-religious private schools in the program. On June 24, 1997, Professor Paul Peterson of Harvard issued a press release describing his team's analysis of test results from these two schools and explaining that "a more extensive examination of the Cleveland School Choice Program is underway to determine if the gains witnessed here are being produced by the entire scholarship program. Results from this evaluation should be available by the fall." Professor Peterson's press release was interpreted by some to mean that his research team was officially evaluating the Cleveland program.

In September 1997, the Harvard Program on Education Policy and Governance (PEPG) report, "An Evaluation of the Cleveland Scholarship Program" was released and drew wide publicity from a *New York Times* article and a *Wall Street Journal* article under Professor Peterson's byline. The report itself was co-authored by Jay Greene (University of Texas, Austin), William Howell (Stanford University), and Paul Peterson (Harvard University).[70]

On December 27, 1997, a front page story in the *New York Times* reported that reading and math scores had improved in both the Cleveland and Milwaukee voucher programs. The only available source of information on test score results in Cleveland was the PEPG report. The *Times* story shows the degree to which the PEPG report is wrongly considered to be the official evaluation of the Cleveland program. In fact, the PEPG report is a privately funded effort that was not commissioned by the Ohio Department of Education.

Although it is titled "An Evaluation of the Cleveland Scholarship Program," the PEPG report describes test score results only from Hope Central Academy and Hope Ohio City Academy. The test results reported are expressed as percentile gains on fall-to-spring testing. It reports overall K-3 percentile gains of 5.6 (reading), -4.5 (language), 11.6 (math total), and 12.8 (math concepts).

The testing regimen whose results are described in the PEPG report was rejected as unsound practice years ago for Federal Chapter I evaluations.[71] Most schools gain every spring and fall back the next autumn. For fall-to-spring changes in test scores to be meaningful, a carefully chosen comparison group must also be tested. The PEPG analysis has no such comparison group. Instead, it makes a comparison to low-income Milwaukee voucher applicants (whose results are not from the same test used by the Hope schools). Therefore, the results reported contribute little to an understanding of how voucher programs might affect student achievement.

Most of the PEPG report details the results of a telephone survey of program applicants. The survey results reported are generally consistent with Witte's findings in Milwaukee that voucher program participants are well satisfied with the program. In the Cleveland survey, parents listed academic quality as their most important reason for participating.

# Vouchers, Values, and Educational Equity

While no strong evidence exists that voucher programs improve student achievement, all parties to the voucher debate at least agree that improving achievement is a desirable goal. But achievement is not the only issue in the debate. People favor or oppose vouchers in part because they hold different social and political values. Professor Peter Cookson (Teacher's College, Columbia University) calls the battle over school choice a struggle over the "soul" of American public education.[72] Jeffrey Henig sees in the struggle a conflict over the type of society Americans want to call into being.[73]

Some observers perceive public schools to have symbolic value as a community institution. In smaller towns, for example, the public high school's athletic teams are community institutions whose support extends beyond the school's students and alumni. In addition, the public character of the school, as expressed, for example, in its availability as a place for meetings, local theater groups, or adult-education programs, contributes to the school's value to the broader community.

Private schools may have considerable symbolic value for their students, parents, and alumni, but rarely for others. By increasing the number and enrollment of private schools, while decreasing those of public schools, large-scale voucher programs would diminish the symbolic value of public schools. In so doing, they could reinforce social fragmentation of the American community along ethnic and racial lines. (This possibility is hinted at by the fact that most Hispanics in Milwaukee went to just one Choice school.)

Large-scale voucher programs may also have the potential to increase inequality and the stratification of students by family income as well as social background. This concern is supported both by theoretical arguments and by empirical evidence on large-scale school-choice programs. (Programs the size of the one in Milwaukee are too small to have much effect on inequality.)

To see how a large-scale voucher program could make school quality and student achievement more unequal, suppose that public schools were replaced by a voucher program.[74] If total spending remained the same as in the public school system, the voucher would be less than the amount formerly spent per student in the public schools because students in private schools who formerly received no public support would now receive a share of this money. For the families of students who previously attended a high-quality private school, the voucher would be equivalent to an increase in income. These families would be likely to spend some of that extra income on better schooling. At the other extreme, students with the lowest level of academic achievement—and whose parents tend to place less priority on education—would receive a voucher lower than the per-student investment within the public school system. The parents of these students would be unlikely to supplement the voucher amount with their own money. If money strongly predicts school quality, these students would, under a voucher system, attend schools inferior to current public schools.

There are two ways to escape the conclusion that vouchers will increase the polarization of educational opportunity. First, if the total investment in public schools increased enough to more than compensate for the spending on students who now attend private school, low-income students might benefit. This seems an unlikely scenario and no current proposal recommends vouchers this large. Second, vouchers might not increase polarization if private schools operated more efficiently than public schools. As we have seen (Box 3), no clear evidence exists that private schools operate more efficiently.

Of course, the current public school system stratifies students by family income and educational background. One of the most important means by which this stratification occurs is residential choice. The more affluent, educated, and committed to education seek to live where their children can attend good schools. The children of the poor are then often left behind to struggle in

substandard, underfunded schools. In his 1995 book, *Private Vouchers*, Terry Moe, one of the most prominent voucher advocates, argues that vouchers are a force for greater educational equity because they provide poor students with a choice of schools.[75] In a voucher system, however, families would sort themselves among schools on the basis of income, educational preferences, and knowledge about schooling. Under the current system, families who send their children to public schools sort themselves among residential locations (and, therefore, school districts) on the basis not only of these factors, but also others, such as the cost and quality of housing, distance to work, and availability of recreational opportunities. For this reason, private schools under a large-scale voucher program are likely to be more internally homogenous (with respect to students' socioeconomic background) than are public schools under the current system. With public schools, some of the poor get a chance to attend the same schools as their middle– and upper–income peers. With large-scale voucher programs, fewer of the poor are likely to have this opportunity. Vouchers would then be a force for educational inequity.

Although not inherent in voucher systems, there are additional features of most voucher proposals that would worsen educational inequity. Most voucher proposals propose considerably lower levels of funding than would result from giving students a per capita share of current spending on education. With this funding, children of affluent parents already in private schools could still spend more than they do currently on education. Children of poor parents would have an even smaller amount to spend on their education. Second, most proposals, including the Milwaukee program, in effect allow private schools to exclude some special needs students, because the schools need not provide services on which those students depend. Some proposals, unlike Milwaukee, would allow schools complete authority over who to admit, and who to exclude. Terry Moe acknowledges the danger that this poses. He argues that it can be addressed through careful attention to the design of voucher programs.[76]

The available empirical evidence supports the contention that vouchers may reduce educational equity. In 1992, the Carnegie Foundation released *School Choice.*[77] Carnegie researchers visited choice programs around the country, surveyed more than 1,000 parents, and reviewed other studies of school choice. Except for Milwaukee's private voucher program, all of the programs in the Carnegie study were public school choice programs. The Carnegie report concluded that:

(1) To the extent that choice programs benefit children at all, they benefit the children of better educated parents,

(2) That the choice programs require additional money to operate,

(3) That choice programs have the potential to widen the gap between rich and poor school districts, and

(4) That school choice does not necessarily improve student achievement.

Bruce Fuller, in a 1995 review of the data available on selected choice programs around the country for the National Conference of State Legislatures, drew conclusions similar to those contained in the Carnegie report.[78]

After a review of the research on school choice in three countries (the U.S., Great Britain, and New Zealand), Geoff Whitty finds little evidence to support the contention that the creation of educational markets increases student achievement. He does, however, find that educational markets make existing inequalities in the provision of education worse.[79] Carnoy draws a similar conclusion based on an analysis of the effects of school privatization in Chile and other countries.[80]

In conclusion, the evidence from Milwaukee and Cleveland reviewed earlier suggests that vouchers have, at best, an uncertain upside. If vouchers could increase educational inequity and social fragmentation, they have a potentially large downside. In this light, Pennsylvania should turn its attention to ideas that have more promise and less danger. One such option, reducing class size, will now be considered.

KEYSTONE RESEARCH CENTER

## BOX 7: DOES MONEY MATTER? SCHOOL SPENDING AND SCHOOL OUTCOMES

Debates about vouchers and class size both touch on a controversial recent debate about whether higher spending improves performance in schools. The holy grail for voucher advocates is improved performance without spending more money. Evidence that money doesn't matter points them to the public education bureaucracy as the problem and to vouchers as a way of achieving better outcomes without necessarily spending more in the long run. Smaller class size, by contrast, would cost more money. The question is whether the performance improvement that results is worth the cost.

University of Rochester Professor Eric Hanushek launched the debate about whether money matters by claiming, based on an extensive analysis of the literature, that "there is no strong or systematic relationship between school expenditures and student performance."[81] The studies Hanushek analyzed attempt to determine the relationship between resource inputs, especially money, and school outcomes. Hanushek's conclusion has been challenged by Hedges, Laine and Greenwald (University of Chicago) based on a meta-analysis of the same studies as Hanushek.[82] Hedges, Laine, and Greenwald find that there is a systematic and educationally important relationship between resources and student achievement. The studies on which both Hanushek and Hedges, Laine, and Greenwald rely have been criticized for being poorly designed, based on nonrepresentative samples, and focused on funding-related characteristics instead of funding as such.[83]

Two other recent strands of literature shed light on the "money matters" debate. While Hanushek's research takes off from the premise that spending on public education has increased rapidly but test scores have not, as noted above, Richard Rothstein's work shows that spending on public education has increased less quickly than generally believed.[84] Moreover, Rothstein estimates that special education spending accounted for 38 percent of net new K-12 spending from 1967 to 1991. The ability of voucher schools in Milwaukee to reject students with exceptional educational needs not only enables the private schools to focus on regular education; it also requires the Milwaukee Public Schools to spend a higher share of funds on special education.

Bruce Biddle (University of Missouri) takes up the question of money in two ways.[85] He uses child poverty data and data on the educational spending of states to study the effects of these two factors on 8th grade math performance. He finds that school funding and child poverty account for 55 percent of the variation in average math achievement among states.

Biddle's findings are in line with results of an earlier study by Ronald Ferguson.[86] Using data from 1986–1990 on 90 percent of the school districts in Texas, Ferguson found that average class size, teacher experience, and the academic ability of teachers accounted for between one quarter and one third of the variation in the reading achievement levels of Texas school districts. He also found that smaller class size and more qualified teachers were more likely to be found in districts that had higher levels of funding.

In a more recent study of fourth and eighth grade math achievement, Harold Wenglinsky (Educational Testing Service) considered how money matters when applied to the funding of school districts.[87] He found that school districts with more students from the least affluent backgrounds have the largest class sizes and are, therefore, least able to raise student achievement. These districts also have the least to spend on central administration. In his analysis, under-funded central administrations ordinarily spend less money on reducing class size and more money on projects with little academic payoff.

Wenglinsky's conclusion that a low pupil-teacher ratio creates a positive classroom social environment and increases math achievement affirms what many parents already appear to know. According to David Figlio and Joe Stone, the higher the pubic school student-teacher ratio in an area, the more likely that parents will send their children to private schools (especially private non-religious schools). Conversely, the higher the private school student-teacher ratio, the more likely parents are to send their children to public schools.[88] This finding suggests much of the debate over the relative merits of public vs. private schools per se may be beside the point.

KEYSTONE RESEARCH CENTER

# SMALL CLASS SIZES

## Historical Background

The impact of class size on achievement has been studied for over a century. Glen Robinson and James Wittebols of the Educational Research Service trace the beginning of research on class size to the work of J.M. Rice in 1893.[1] In a 1902 study, Rice concluded that there was no relationship between class size and student achievement.[2] In subsequent decades, the heyday of the industrial model of schooling, much research on class size aimed at ascertaining how large classes could be made without significant reductions in student learning.

Howard Blake's 1954 review of 267 different studies marks the beginning of modern class size research. Of the 85 original studies Blake found that focused on elementary and secondary education, 22 met his criteria for qualifying as scientific studies. Of these 22, 16 found that children learned more in smaller classes, three favored larger classes, and three were inconclusive.[3]

During the 1960s and 1970s, attention turned to the impact that small group instruction might have on children from low-income families.[4] Many studies in this period statistically explored whether Chapter 1 funding improved the performance of low-income students relative to comparable or more advantaged ones who received no support. This research left the question of whether low-income children benefit from smaller classes unanswered.

In 1978, Professor Eugene Glass of Arizona State and Mary Lee Smith published an influential and controversial meta-analysis of studies conducted in more than a dozen countries.[5] Glass and Smith concluded that small classes produce higher levels of student achievement than large classes. For example, they found that being taught in a one-on-one tutorial as opposed to a 40–student class improved student performance by 30 percentile ranks. Glass and Smith argued that to be most effective classes should have about 15 students.

Robinson and Wittebols argued that Glass and Smith had drawn conclusions based on too few studies and that they relied too much on research on individual tutoring.[6] Professor Robert Slavin of Johns Hopkins considered Glass and Smith's analysis flawed because it did not carefully enough take into account qualitative distinctions between studies.[7] In Slavin's view, except for studies of class sizes of one, Glass and Smith's evidence that class size reductions raised achievement was weak.

### BOX 8:
### LOW PUPIL-TEACHER RATIOS DO NOT ALWAYS MEAN SMALL CLASS SIZES

The terms pupil-teacher ratio and class size are often used interchangeably in everyday conversation. Most people understand both terms to mean the average number of students in a typical classroom with one teacher. This is a false assumption. In 1996, for example, the average elementary school pupil-teacher ratio in the U.S. was 18.8:1 and the average secondary school pupil-teacher ratio was 14.7:1.[8] In 1993–94, the average class size in "self-contained" public school classrooms (in which students are taught primarily in one room by one teacher, as in most elementary schools) was 25.2 students. Classrooms in departmentalized schools (in which students move from class to class, being taught by different teachers) had enrolled 23.2 students on average.[9]

One calculates pupil-teacher ratio by dividing the number of students by the number of instructors holding teaching certificates whose primary responsibility it is to teach. These instructors include teaching specialists in areas such as physical education, art, reading, and special education, as well as Chapter I "pull out" teachers (pull-out teachers remove students from the regular classroom who qualify for means-tested specialized instruction.) One calculates average class size by surveying classroom teachers and asking how many students are in their classes. Average class size is a better indicator of the overall classroom experience of most teachers and most students than is the pupil-teacher ratio.

The Pennsylvania and U.S. pupil-teacher ratios are very similar—17.3:1 for the United States compared to 17.1:1 for Pennsylvania in fall 1994.[10] In 1993, large U.S. central cities had pupil-teacher ratios of 19.0 compared to the overall U.S. average that year of 17.8. Medium-sized central cities had pupil-teacher ratios of 17.9:1. Urban fringe areas had pupil-teacher ratios of 18.3:1 to 18.6:1. While urban pupil-teacher ratios are only a little above average, urban class sizes may be more substantially above average. Indirect evidence for this comes from the research of Professor Michael Boozer of Yale and Professor Cecilia Rouse of Princeton.[11] Boozer and Rouse use four data sources: responses to a telephone survey of a random sample of 500 New Jersey teachers, information on New Jersey schools from the state Department of Education, and two national data bases. In all four sources, Boozer and Rouse find that pupil-teacher ratios in schools with high percentages of African-Americans are not significantly different from ratios in mostly white schools. The New Jersey survey and national data base with information on class size, however, show that heavily black schools have significantly larger class sizes – there are an estimated three or four children more per class in a hypothetical all-black as opposed to all-white school. Within each class type—e.g., regular, gifted, or special needs—blacks also attend larger classes.

Boozer and Rouse report that smaller eighth grade class sizes lead to larger test score gains by 10th grade. Differences in class size explained about 15 percent of the difference in the black-white achievement gain between eighth and 10th grade.

Boozer and Rouse's findings illustrate the importance of targeting reduction at actual class size for students in regular classes in urban schools.

## Indiana Prime Time

Against this backdrop of controversy over the relationship between class size and student achievement, Indiana launched Project Prime Time, a state-wide class size reduction effort.[12] In 1984, Indiana school corporations (the Indiana equivalent of school districts) became eligible to receive state funds to pay the salaries of additional teachers and teacher-aides that are necessary to reduce corporation-wide first grade class size averages to 18, or to 24 if a teacher-aide was in the room. In 1985, the state extended this arrangement to second grade and in 1986 corporations gained the option of adding either kindergarten or third grade. Now in its 14th year, Prime Time today subsidizes the salaries needed to move toward corporation-wide average class targets of 18 students per teacher in K-1 and 20 in grades two and three. In recent years, Prime Time has been accompanied by an extensive effort to provide professional development and disseminate instructional methods that take full advantage of small class sizes.

Research on Prime Time showed mixed results. In 1990, David Gilman and Christopher Tillitski, after reviewing four studies, concluded that Prime Time class size reductions had produced no achievement advantage.[13] They cautioned that their findings did not necessarily imply that any class size reduction program would fail. Prime Time was, in their judgement, a poorly conceived, hastily implemented program with inadequate provision for training teachers and for systematic evaluation.

State-funded evaluations of Prime Time, conducted in 1987 and 1992, showed positive but not definitive results.[14] The 1992 evaluation examined the experiences and test scores of 21 schools in 12 districts, but did not include a control group.[15] The evaluation found that, after two consecutive years in Prime Time, third grade students outscored the state-wide average student on the Indiana State Test of Educational Process (ISTEP), a battery of language, math, and reading tests.[16] Students in small classes in grades one and two beat the state-wide average by more than students in small classes in grades two and three. Sixth graders who had Prime Time in first and second grade did better on the ISTEP than the state-wide average, but sixth graders who had Prime Time in grades two and three did not, possibly because they were drawn mostly from large city and poor rural districts.[17]

KEYSTONE RESEARCH CENTER

# The Tennessee Experience

## The Tennessee Star Study

The Tennessee STAR experiment is exactly the kind of carefully designed approach to studying the effects of class size reductions called for by Gilman and Tillitski. In the mid-1980s, the Tennessee legislature became interested in the possibility that reducing class size could increase student achievement. Key legislators knew of Indiana's Prime Time program and a class size study conducted in Nashville[18], as well as the research literature. They were particularly influenced by the meta-analysis done by Glass and Smith, which suggested reducing class size to about 15. Mindful of the cost of reducing class size, the legislature wanted to study the impact of reducing class size in the early grades before adopting a class-size reduction policy.

In 1985, the Tennessee legislature passed, and governor Lamar Alexander signed into law, funding for a state-wide class size experiment. The Student/Teacher Achievement Ratio (STAR) study followed a group of students from kindergarten through third grade. Since Tennessee did not require kindergarten, many STAR students entered the study as first graders.

A consortium of researchers from Memphis State University, Tennessee State University, the University of Tennessee at Knoxville, and Vanderbilt University carried out the STAR study. The state appropriated $9 million over four years to pay for the additional teachers and teacher aides necessary to reduce class sizes in selected schools, and $3 million to support the study itself.

The STAR study began in the fall of 1985 in 79 schools within 42 school districts throughout the state.

Researchers classified schools as:

1) inner-city (metropolitan-area schools in which more than half the students received free or reduced-price lunches);

2) urban (schools in towns of more than 2,500 serving an "urban" population);

3) suburban (districts located in a metropolitan area's outer fringe), and

4) rural.

Within each participating school, the State Department of Education randomly assigned teachers and students to one of three types of classes: small (S) classes (13-17 students),

### BOX 9: KEY FINDINGS FROM ANALYSES OF TENNESSEE CLASS SIZE DATA

1. Statistically significant differences were found between small classes and the two types of regular classes on every achievement measure in every year of the study.
2. The small-class advantage was greatest in the first year that the student entered a small class, whether kindergarten or first grade, and remained stable through second and third grade.
3. Achievement benefits of small classes in K-3 continued through at least grade eight.
4. In each grade, minorities and students attending inner-city schools enjoyed greater small-class advantages than whites on some or all measures.
5. In grades one to three, all students benefited significantly when a high proportion of their classmates had attended kindergarten.
6. Students in small classes had higher test scores in a wide range of subjects, establishing a solid foundation for a rich life and a rich variety of future careers.
7. The same benefits from small classes were found for boys and girls alike.
8. Every type of district—inner city, urban, suburban, and rural—enjoyed significant gains from small classes.

(S) classes (13-17 students), regular (R ) classes (22-25 students), and regular classes with a full-time instructional aide (RA) (22-25 students).

With one important exception, once assigned to a class type, students stayed in that type of class as long as they remained in STAR. The major exception was that students in regular and regular with aide classes during kindergarten were randomly reassigned to either R or RA classes for first grade. (The researchers observed no significant differences between R and RA student performance in kindergarten.) While it complicates analysis of the STAR experiment, the reassignment does not interfere with the central findings regarding the performance of students in small classes versus the other two types.

To insure that curriculum differences, leadership style, school climate, and other school-specific factors did not influence the results, all schools participating in the project had to be large enough to have all three types of classes at all four grade levels. The STAR project also dictated that there be no changes in participating schools other than the establishment of the three types of classes.

In sum, STAR was one of the few truly randomized experiments ever conducted in education. It was also a large and well-designed study. The project began with about 6,000 students and by its conclusion had approximately 11,000 pupil records in its data base. In the STAR longitudinal data base for K-3 there are 54 schools, 207 classrooms, and 1,842 students. For K-1 there are 74 schools, 307 classrooms, and 2,416 students. For first to third grade there are 60 schools, 236 classrooms, and 2,571 students.

Students in STAR were tested in reading and math on the (nationally normed) Stanford Achievement Test (SAT) and the (state criterion-referenced) Tennessee Basic Skills First (BSF) test. STAR researchers compared improvements in achievement each year by each class type. They also compared the performance of students in small classes for three consecutive years with the performance of students in each type of regular class for three consecutive years.

STAR researchers found that students in small classes out-performed students in both R and RA classes across the board in all geographical areas and at all grade levels. Regular classrooms with a teacher aide did show a slight but not statistically significant achievement advantage over regular classrooms in first grade. Results for classes with an aide were otherwise mixed and somewhat contradictory

Averaged over four years, students in small classes had an advantage of a bit more than 8 percentile ranks over students in regular classes in reading and a bit less than 8 percentile ranks in math (Figures 5 and 6). The effect size (see Box 2) in reading averaged over four years is about .26. In math it is .23. Students who started in small classes in kindergarten established an achievement advantage in their first year and then maintained it during the next three years.[19]

In a May 1997 reexamination of the STAR data, economist Alan Krueger of Princeton confirmed the original findings of the STAR investigators.[20] Krueger controls for other measured factors that might influence performance, including student characteristics (race, gender, eligibility for free lunch, whether the student was new to the school, etc.) and teacher characteristics (race, gender, experience, and educational qualifications). Given the original random allocation of students and teachers, these characteristics should not influence the impact of class size on performance. As expected, Krueger finds that controlling for these variables has very little effect. Krueger still finds overall effect sizes that range from 0.19 to 0.28 in the four years—similar to the range reported in the original STAR analysis.[21]

In a sample containing students in all grades, Krueger finds that the achievement of students in small classes jumps by about 4 percentile ranks in the first year a student attends a small class and improves by almost an additional percentile point for each additional year. The initial effect is highly significant, while the incremental improvement in subsequent years is just on the margin of statistical significance. Krueger also shows that having a high proportion of classmates who attended kindergarten has a large, positive impact on individual achievement.

The original STAR results may be understated because some classes labeled as small were actually larger than some labeled as large. (Since the number of students in a grade does not fall into multiples of 15 and 23-24, it is unavoidable that small and regular classes be distributed around these targets.) A research

**Figure 5: Math Achievement Gains for Tennessee Students in Small vs Regular Classes, by grade**

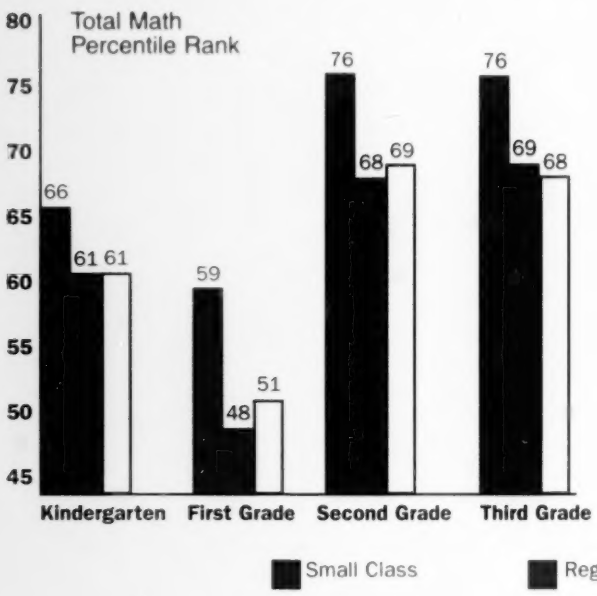
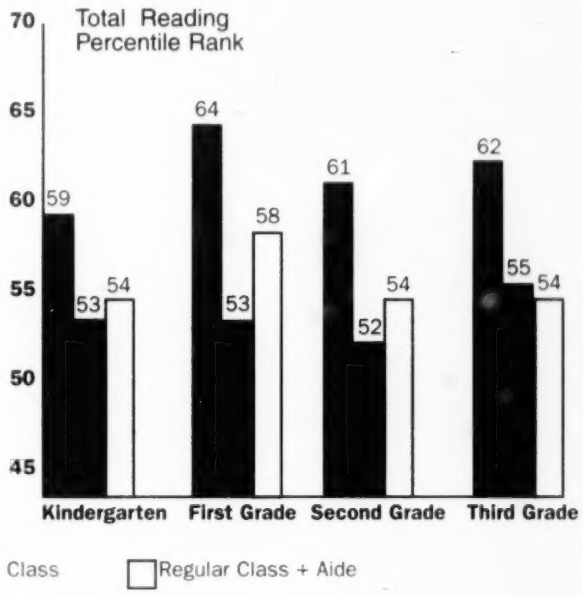



**Figure 6: Reading Achievement Gains for Tennessee Students in Small vs Regular Classes, by grade**




Small Class  Regular Class  Regular Class + Aide

Source: Elizabeth Word, Charles M. Achilles, Helen Bain, John Folger, John Johnston and Nan Lintz, "Project STAR Final Executive Summary Report" (Nashville, TN: Tennessee State Department of Education, June 1990).

team headed by Professor Barbara Nye and B. DeWayne (Tennessee State University) reestimated the performance difference of small classes and regular classes after removing all small classes that did not have 12-14 students and all regular classes that did not have at least 23 students from the sample. They report effect size advantages for small classes that average .56 for reading and .47 for math. Further, some of the effect sizes increase from first through third grade.[22]

The STAR study also found that small classes especially raised achievement in inner-city Tennessee classrooms with large concentrations of minority students.[23] (See Figures 1 and 2 in the Executive Summary.) Jeremy Finn (State University of New York at Buffalo), and Charles Achilles (now at Eastern Michigan University)—both now consultants to Tennessee State University—report that, for minority students, an eight-student reduction in class size resulted in achievement gains of 0.35 of a standard deviation (i.e., an effect size of 0.35) in reading and 0.23 in math. This compares with gains of 0.13 and 0.15 for whites.[24]

On first grade tests, the gains made by minority students as a result of attending kindergarten were twice as large as those made by white students. Achilles and Finn also found that, on the Basic Skills First (BSF) reading test, the difference in the pass rate between white and minority students was reduced from 14.3 percent in regular classes to 4.1 percent in small classes. The same pattern was repeated for word study skills and math, although not at statistically significant levels. Krueger also finds that lower achieving, minority, and poor students benefit the most from attending smaller classes.[25]

Charles Achilles, Jeremy Finn, and Helen Bain report that when both white and non-white Tennessee students began kindergarten in small classes, 87 percent of white and 86 percent of non-white first graders passed the Basic Skills First test. For students who began kindergarten in regular classes, the non-white first grade pass rates trailed white by 12 percent.[26]

Steven Bingham, after conducting a review of the research literature on the white-black

test gap and on class size, including the Tennessee experience, concluded that small class size in the early grades is an effective achievement gap reduction strategy. He maintains that minority children should be placed in small classes early (preferably kindergarten) and remain in a small class for at least two years.[27]

The STAR study found that small classes increased promotion rates from each grade. Over the four years of the study, 80.2 percent of students in small classes moved up to the next grade the following year, compared with 72.6 percent of students in regular classes. Raising promotion rates for each grade saves money by reducing the number of students taught twice at each grade level.[28]

In addition, when more students are held back, the R and RA classes at the next grade level end up with fewer low-scoring students. If students in R and RA classes had been promoted at the same rate as in small classes, the relative test scores of R and RA classes might have been even lower. The higher-retention-in-grade rates of R and RA classes may bias downward the estimate of the additional benefit of several years in a small class.

Finally, the Tennessee experiment provides some evidence that small classes mitigate the negative effect of large schools documented by William Fowler and Herbert Walberg (University of Illinois at Chicago).[29] According to Achilles, students in regular classes achieved less well in large schools than small schools. Students in small classes did as well or nearly as well in large schools as they did in small schools.[30]

Given its scope, its careful randomized experimental design, and the power of its data, Harvard Professor Frederick Mosteller, in a report to the American Academy of Arts and Sciences, characterized the STAR study as "one of the great experiments in education in United States history."[31]

## The Tennessee Lasting Benefits (LBS) Study

The STAR experiment was followed in 1989 by the Lasting Benefits Study (LBS), coordinated by Barbara Nye at Tennessee State University. The Lasting Benefits Study tracks STAR students as they continue their school careers. A STAR student is defined in the LBS as any student who spent at least third grade in a STAR classroom. This means that, in the effect sizes below, students who only spent grade three or grades two and three in small classes are included. Including these students makes the following estimates of the long-run impact of small K-3 classes conservative.[32]

At least through eighth grade, students in small classes during K-3 continue to perform better academically than graduates of R and RA classes. This achievement difference is still statistically significant.[33] The achievement advantage for minority students who participated in small classes remains larger than that for white students.[34]

Results from the Lasting Benefits Study show eighth-grade effect sizes of 0.04 to 0.08,[35] seventh-grade effect sizes that range from .08 to 0.16,[36] sixth grade effect sizes that range from .14 to .26,[37] fifth grade results ranging from .17 to .34,[38] and fourth grade effect sizes of .11 to .16.[39] While STAR students from small classes continue to outperform students in regular classes, the presence of a teacher-aide continues to have very little, if any, impact on achievement.

The LBS reports that lasting benefits from K-3 small class sizes result for a wide spectrum of subjects, including reading, language, math, study skills, science, and social studies.[40]

## Project Challenge

Beginning in 1989, Project Challenge provided the money necessary to reduce K-3 class size in 16 of Tennessee's poorest school districts. These districts typically placed low on achievement rankings of Tennessee's 138 school districts. Since the implementation of Project Challenge, student achievement in math and reading has improved both in comparison to the performance of previous students in these districts and in relation to other schools in the state.[41] Between 1989–90 and 1993–94, the average ranking on grade two test results of Project Challenge school districts improved from 97 to 78 in reading and from 90 to 56 in math.[42] In other words, student achievement in these poor districts was only a little below the median district in the state in 1993–94 and above the median in math.

KEYSTONE RESEARCH CENTER

# Nevada and California

## Nevada

Nevada passed its Class Size Reduction Act in 1989 and implemented it in first grade and selected, at-risk kindergartens in the 1990–91 school year. Second grade was added in 1991–92 and third grade partially implemented in 1996–97.[43] Only 60-70 percent of the first and second grade classes in the Nevada program reduce class size by establishing a classroom with one teacher and 15 students. The rest use flexible groupings, multi-age grouping, two teachers with 30 students sharing a classroom, etc.

According to Dr. Mary Snow of the State Department of Education, Nevada has never devoted the resources necessary to conduct a full-scale systematic evaluation of its class-size initiative.[44] Dr. Snow published evaluations of the Class Size Reduction Program in 1993 and 1997. James Pollard and Kim Yap of the Northwest Regional Educational Laboratory prepared an evaluation in 1995.

Snow's 1997 evaluation shows that having attended small classes in earlier grades significantly improves mean test scores in language, math, and reading for fourth graders.[45] Improvements in scores were generally small, especially for reading.[46] The 1995 evaluation found that, in parts of the state, students in larger classes actually scored better in reading.[47]

The mean scores for students with lower socio-economic status and for minority students do not show differences based on participation in the Nevada Class Size Reduction Program. Results from Nevada generally favor teaching in self-contained classes as opposed to team teaching in rooms of about 30 students.

In a 1995 Nevada opinion survey, 61 percent of parents believed that the benefits warranted paying the estimated $852 per student for smaller classes. Less than 10 percent of parents believed the benefits were not worth the cost.[48]

## California

Beginning in the 1996-97 school year, California began implementing an ambitious class-size reduction program.[49] In the first year, districts that reduced class sizes to 20 or below received $650 for each student enrolled in a class of no more than 20 students. The 1997 California budget raised the allotment to $800 per student and contained almost $1.5 billion for class size reduction. Schools must start by reducing first grade class size, then second grade, and then either kindergarten or third grade.

In the first year, 18,400 new teachers were hired to implement class size reduction in California. Moreover, California already had the nation's fastest growing student enrollment. One consequence is that 30 percent of newly hired teachers state wide were uncredentialed in the first year of the California class-size reduction program. Two-thirds of those hired in Los Angeles do not possess teaching credentials.[50]

Despite teacher and facilities shortages, teacher and parent response has been overwhelmingly positive. In Stanislaus County, for example, a survey conducted with the assistance of the San Diego County Office of Education found that 76 percent of parents and 96 percent of teachers felt that the reading skills of students in smaller classes were much or somewhat improved. Eighty-nine percent of parents said the benefits were worth the $1 billion-plus that the state and local schools spent to reduce class sizes in the first year.[51] Asked the same question, 97 percent of parents in Coronado County said the program was worth the state's investment.[52]

There are several reasons to question whether the California initiative will show the test gains seen in Tennessee: the selection of 20, not 15 as a target class size; the inexperience of new teachers; facilities crowding; and the limited amount of training received in how to make small classes effective. So far, no performance evaluation of the California class-size reduction program has been put in place.[53] California did not put a systematic evaluation program in place from the beginning. No baseline data were collected prior to small class size introduction. The state has not yet funded the mandated evaluation called for by the year 2002. A consortium of research organizations in

concert with school districts and associations is planning a multi-year comprehensive study.[54] The aim is to encourage information-sharing and learning by practitioners as well as to add to the research literature. The initial research design will focus on successive cohorts of third and fourth graders who have and have not attended smaller classes.

## BOX 10: MILWAUKEE'S PUBLIC SCHOOLS WITH SMALL CLASSES OUTPERFORM CHOICE SCHOOLS

As Cecilia Rouse noted in her first analysis of the Milwaukee voucher program, the Milwaukee data set lacks any information on school or class variables. This makes it difficult to explain any difference between the test scores of comparable students at the "average" Milwaukee Public School and the "average" Choice school. Do scores differ because of the inherent differences between public and private schools, or because of some variable (such as class size) that happens to coincide with private or public status?

Having raised the question of what actually takes place inside Milwaukee's public and private schools, Rouse, in a subsequent paper, takes a first step toward actually opening the "black box" to take a look.[55] She does this by first observing that at least three distinguishable sub-groups of schools exist within the 145 schools of the Milwaukee Public School District. The district includes about 30 magnet schools created in the 1970s to promote desegregation; these draw their students from throughout the city. Magnet schools enroll about 22 percent of the total MPS enrollment. In addition to magnet and regular schools, the Milwaukee district includes 14 schools that were exempted from desegregation in the 1970s and provided with extra funding from the state. Today, these 14 schools, and 9 others, are known as P-5 schools. P-5 schools enroll about 15 percent of the total public school students and 25 percent of the elementary school children. To remain eligible for state grants to schools with high proportions of economically disadvantaged and low-achieving students, P-5 schools must maintain pupil-teacher ratios of 25 or less. They must also meet a variety of other conditions—including conducting annual testing in basic skills, increasing parental involvement, and identifying students needing remedial education. Rouse, however, regards small classes as the most distinctive feature of P-5 schools. A state allocation of $6.7 million allows about $500 for each child in P-5 schools.

Rouse examines the test scores of students in regular schools, magnet schools, and P-5 schools. Results that do not adjust for family background and student ability show that students in the magnet schools consistently score better than students in the regular public schools. The gap increases the longer students attend the magnet schools. Students in P-5 schools and voucher recipients have lower scores than magnet school students, but the difference does not increase over time. Once family and student characteristics have been accounted for, the gap in math scores between the magnet and regular schools disappears. The gap in math scores between lower-achieving magnet and regular schools, on the one hand, and higher-achieving P-5 and Choice schools, on the other, becomes large and statistically significant. For reading, controlling for background characteristics, students in the P-5 schools have faster gains than any other group, including voucher students.

What explains the test scores of these sub-groups, including the high performance of the P-5 schools? Rouse shows that the average pupil-teacher ratio in P-5 schools is 17:1, compared to between 19:1 and 20:1 at magnet schools and at regular MPS schools. Five Choice schools that she contacted by telephone have a 15.3:1 pupil-teacher ratio—lower even than P-5 schools. The Choice relative class size might be even smaller than its pupil-teacher ratio because Choice schools have fewer special education responsibilities.

Rouse concludes that smaller class size could explain both the Choice and P-5 advantage in math. Small class size does not explain the advantage in reading that P-5 schools enjoy over Choice schools. To explain that would require shining more light on the black box, and finding out what other features make P-5 public schools more effective at teaching reading.

KEYSTONE RESEARCH CENTER

# Wisconsin: Student Achievement Guarantee In Education (SAGE)

Wisconsin implemented its statewide Student Achievement Guarantee in Education (SAGE) program in 1996–97. SAGE seeks to increase the academic achievement of children living in poverty by reducing the student-teacher ratio in kindergarten through third grade to 15:1.[56] Participation in SAGE requires a school to implement a rigorous academic curriculum, provide before–and–after school activities for students and community members, and implement professional development and accountability plans.

Any district with a school that enrolls 50 percent or more low-income children in a school could participate. Within eligible districts, any school enrolling 30 percent or more low-income children could apply. Each district, except Milwaukee, could designate one school as a SAGE school. Milwaukee was allowed 10 SAGE schools.

Schools entering the program had to agree to remain in SAGE for its five-year duration and they also had to submit an annual "Achievement Guarantee Contract" to the Department of Public Instruction. This contract explains how the school plans to implement the SAGE program requirements. Schools are allowed wide latitude in developing their plans. Upon accepting a school into SAGE, the state provides up to an additional $2,000 per low-income student enrolled in SAGE classrooms. While the original legislation specified that no new schools would be admitted after the start of the 1996–97 school year, SAGE proved so popular that the state legislature agreed to expand it beginning with the 1998–99 school year.

SAGE is designed to be implemented in stages. Kindergarten and first grade classes entered the program in 1996–97, second grade was added in 1997–98, and third grade will be added in 1998–99. All classrooms at the appropriate grade level in participating schools must have a pupil-teacher ratio of no more than 15:1. During the 1996–97 school year, SAGE was implemented in 30 schools (seven in Milwaukee) throughout Wisconsin. It encompassed 84 kindergarten classrooms, 96 first–grade classrooms, and 5 mixed–grade classrooms. SAGE classrooms enrolled 1,715 kindergarten and 1,899 first–grade students.

The legislation creating SAGE requires an annual evaluation of the program and a fifth–year final report on the impact of the program on academic achievement. This legislatively mandated evaluation is being conducted by Alex Molnar and co-authors at the Center for Urban Initiatives and Research at the University of Wisconsin-Milwaukee. SAGE schools are being compared to a group of 16 non-SAGE schools in SAGE districts. Comparison schools were selected for their similarity to one or more individual SAGE schools in demographic composition, school size, third–grade test scores (initially), and percentage of low–income students. In addition to quantitative analysis, the SAGE research methodology includes an extensive and systemic protocol of qualitative research, including interviews of teachers and principals, surveys of teachers, teacher logs, and classroom observation.

In the first annual evaluation of the SAGE program, released in December 1997, Molnar and co-authors compared the academic performance of students in SAGE first–grade classrooms to that of students in comparison–school first grade classrooms using a "before" test in October 1996 and an "after" test in May 1997.[57]

October 1996 results showed no statistically significant differences between SAGE and comparison–group student performance. In May 1997, SAGE students scored significantly higher in reading, language arts, and math. Overall, the achievement gain for SAGE students equaled 12–14 percent more than the gain for the comparison group. (See Table 3.)

Controlling for pretest score, subsidized lunch eligibility, days absent, and race, small class students scored significantly above their non-SAGE counterparts on every test.

African-American males, in particular, appear to benefit from participation in the SAGE program. The total scores of African-American males on all three tests rose 56 points in SAGE classrooms compared to 39.4 for the matched schools. (See Figure 3 in the Executive Summary.) As a group, African-American students scored lower than white students on the October test in both SAGE and comparison-group schools. May results show that the gap between the achievement of African-American students, as a group, and white students, as a group, widened in comparison-school classrooms. In contrast, in SAGE classrooms, African-American students, as a group, and white students, as a group, increased their achievement by similar amounts. It should be borne in mind that these results are for the first year of the program. Therefore, SAGE first-graders tested had not attended SAGE kindergartens. Also, it is possible that a number of SAGE and comparison school first-graders did not attend kindergarten at all or attended half-day programs.

In their qualitative research, evaluators found that in SAGE classrooms:

1) little time is required to manage the class or to deal with discipline problems;

2) much time is spent on instruction, actively teaching;

3) a large portion of instruction is individualized and spent in diagnosing student needs, providing help, and in monitoring progress; and

4) students showed increases in "on task" and "active learning" behaviors over the year. These behaviors were also found to be related to SAGE student performance on the CTBS.

In general, the first year SAGE results appear to be tracking the results of the STAR study.

**Table 3: Change in Mean Test Scores from October 1996 to May 1997—First Graders in Small Classes and in Regular Classes**

| | Change in Mean Test Score from October 1996 to May 1997 | | | |
|---|---|---|---|---|
| Test | SAGE Schools: Small Classes | Comparison Schools: Regular Classes | SAGE Advantage | % Difference in SAGE Increase in Score |
| Language Arts | 53.8 | 46.2 | 7.6 | 14% |
| Reading | 51.3 | 45.0 | 6.3 | 12% |
| Mathematics | 55.4 | 48.5 | 6.9 | 12% |
| Total | 53.4 | 46.5 | 6.9 | 13% |

Source: Peter Maier, Alex Molnar, Philip Smith, John Zahorik, *First-year Results of the Student Achievement Guaranteee in Education Program* (Milwaukee: Center for Urban Initiatives and Research, University of Wisconsin-Milwaukee, December 1997), Table 23.

KEYSTONE RESEARCH CENTER

# RECOMMENDATIONS

There is no longer any argument about whether or not reducing class size in the primary grades increases student achievement. The research evidence is quite clear: it does.

One of the most powerful illustrations of the impact of smaller class sizes comes from research by Harold Wenglinsky on 203 school districts.[1] On the 1992 National Assessment of Educational Progress Mathematics test, Wenglinsky found that fourth graders in smaller-than-average classes were about four months ahead of fourth graders in larger-than-average classes. In a sub-group of primarily large, urban schools, fourth graders in smaller-than-average classes were three-quarters of a school year ahead of their counterparts in larger-than-average classes.

In contrast, the claim that participation in a voucher program increases student achievement is weak. It rests almost entirely on analyses of data from the Milwaukee Parental Choice program. The number of students is small and the data sets often fragmentary. Using a variety of statistical techniques, two of the three analyses (Witte and Rouse) find no achievement advantage for Choice students in reading. Two analyses (Greene, Peterson, and Du, and Rouse) find a modest achievement advantage in math for choice students. However, these results are derived by applying a variety of complex and sometimes controversial analytic methods to weak data. As Cecilia Rouse cautions, data limitations threaten the validity of any evaluation of the Milwaukee Parental Choice Program. As she points out, the econometric techniques she deployed in her analysis of the Milwaukee program can not substitute for better data.

Some may suggest that the inconclusiveness of the Milwaukee results argues in favor of further voucher experiments to be implemented so the idea can be tested further. Such a suggestion might have merit if there were no clearly superior strategies for promoting the academic achievement of low-income students. As it stands, there is strong, clear, and consistent evidence that reducing class size to 15 in kindergarten and first grade significantly improves academic achievement. Moreover, the results of additional voucher experiments on a small scale cannot be generalized to produce conclusions about the

## BOX 11: WHY ARE SMALL CLASS SIZES SO EFFECTIVE?

The SAGE, STAR, and other studies reviewed in this report suggest that small classes promote higher achievement for a range of mutually reinforcing reasons.

- Children receive more individualized instruction.
- Teachers can focus more on direct instruction and less on classroom management.
- Students become more actively engaged in learning than peers in large classes.
- Teachers identify learning disabilities sooner, but fewer children end up going into special education classes because teachers can support them within small classes.
- Teachers are more able to give children from low-income families and communities a critical, supportive adult influence.
- Teachers are better able to engage family members and to work with parents to further a child's education.
- Teachers of small classes less often burn out.

likely impact of a large-scale municipal or state voucher program. Small scale class-size reduction experiments, in contrast, do tell us what to expect from across-the-board class-size reduction.

Policy makers considering education reforms to improve the achievement of low-income children should carefully consider the strength of the evidence and the quality of the research on smaller class sizes. In policy making there is sometimes a tendency to regard all studies and research reports as being created equal. They are not. As Princeton economist Alan Krueger put it, referring to the STAR study, "One well designed experiment should trump a phalanx of poorly controlled, imprecise observational studies based on uncertain statistical specifications."[2]

The scholarly discussion about the academic impact of class size reductions is settled as far as whether they generate benefits. What remains is a discussion about: 1) whether the achievement gained is worth the cost; 2) whether the class size reductions should be general or targeted; and 3) how class size reductions should be used in conjunction with other academic strategies.[3]

Despite some disagreement, there is a strong consensus that targeting class size reductions on kindergarten and first grade will provide the greatest academic gains for the money invested. It is also widely agreed that reducing class size is a preventive strategy, not a remedial strategy. In other words, children should be taught in small classes at the earliest possible point in their school careers and reductions in class size should be used as a base upon which additional educational strategies are built. Thus, small classes in kindergarten and first grade should be seen as a strong foundation for other strategies such as "Success for All" and "Reading Recovery" which have had good results increasing the reading achievement of low-income children.

Since the evidence indicates that small classes generate the greatest gains in kindergarten and first grade, this report recommends that Pennsylvania:

1) Provide universal, publicly funded full-day kindergarten with student-teacher ratios of 15:1; and

2) Reduce class size in first grade to 15.

Research suggests that more modest gains result from small classes in grades two and three. In addition, considerable scope for innovation exists in exploring how to build on gains established in small kindergarten and first grade classes. Therefore, this report recommends that Pennsylvania:

3) Implement an experimental program in which class size reductions for grades two and three are achieved in a variety of ways.

To make for a smooth transition and avoid teacher and classroom shortages of the kind observed in California, these recommendations should be phased in over time. Implementation should be targeted initially at the schools and communities most in need—those in the bottom quarter of schools, measured by family income and test scores. Implementation in these schools should begin with kindergarten in the first year and first grade in the second year. The experimental program of class-size reductions in grades 2 and 3 should begin in the third year. Scaling up class-size reduction in grades 2 and 3 can be done once we know the best ways to add to the gains achieved in grades K-1.

Small class sizes and all-day kindergarten should be implemented systematically. Researchers should collaborate with policymakers and practitioners so that lessons learned in the early stages allow for cost-effective implementation of small classes for all K-3 students in the state. For grades 2–3 and in schools that miss the initial implementation cut-offs, the research design could include some controlled within-school experiments along the lines of the Tenneessee STAR experiment.

Pennsylvania could implement these recommendations by making an investment of roughly $100 million in each of the first two years.[4] This is a small fraction of Pennsylvania's projected budget surplus for 1997-98. This amounts to an investment of about $8.33 each for the state's 12 million residents. Pennsylvania's children are worth this investment.

KEYSTONE RESEARCH CENTER

# REFERENCES

**Vouchers**


1 W. Van Vliet and J. A. Smyth, "A Nineteenth-Century French Proposal to Use School Vouchers," Comparative Education Review 26 (February 1982): 95-103.

2 National Commission on Excellence in Education, A Nation at Risk: The Imperatives for Educational Reform (Washington, D.C.: U.S. Department of Education, 1983).

3 Milton Friedman, Capitalism and Freedom (Chicago: University of Chicago Press, 1963).

4 Jeffrey R. Henig, Rethinking School Choice: Limits of the Market Metaphor (Princeton, N. J.: Princeton University Press, 1993), p. 104.

5 David Tyack, The One Best System: A History of American Urban Education (Cambridge: Harvard University Press, 1974).

6 Ivan Illich, Deschooling Society (New York: Harper and Row, 1971).

7 Allen Graubard, Free the Children: Radical Reform and the Free School Movement (New York: Pantheon, 1972). See also the letter from Herb Kohl to Mario Fantini printed in the appendix to, Mario D. Fantini, Free Schools of Choice (New York: Simon and Schuster, 1973).

8 Paul Goodman, The New Reformation: Notes of a Neolithic Conservative (New York: Random House, 1970).

9 Education Vouchers: A Report on Financing Elementary Education by Grants to Parents (Cambridge, MA: Center for the Study of Public Policy, 1970), as cited in Richard F. Elmore, Choice in Public Education (Santa Monica, CA: The Rand Corporation, 1986), p. 9.

10 Amy Stuart Wells, Time to Choose: America at the Crossroads of School Choice Policy (New York: Hill and Wang, 1993), p. 152.

11 Wells, Time to Choose, p. 174.

12 Charles J. Russo and Michael P. Orsi, "The Supreme Court and the Breachable Wall," Momentum, September 1992, pp. 42 - 45.

13 Thomas W. Lyons, "Parochiaid? Yes!" Educational Leadership, November 1971, pp. 102-104, and Glenn L. Archer, "Parochiaid? No!" Educational Leadership, November 1971, pp. 105-107. See also Grace Graham, "Can the Public School Survive Another Ten Years?" Educational Leadership, May 1970, pp. 800-803.

14 "Justice and Excellence: The Case for Choice in Chapter 1," U.S. Department of Education, November 15, 1985, as cited in Elmore, Choice in Public Education.

15 Henig, Rethinking School Choice.

16 Henig, Rethinking School Choice, chap. 4.

17 Education Commission of the States, "Legislative Activities Involving Open Enrollment (Choice)," Clearinghouse Notes, December 1994.

18 Education Commission of the States, "Legislative Activities Involving Open Enrollment (Choice)," p. 91.

19 Jeffrey R. Henig, Rethinking School Choice, p. 67.

KEYSTONE RESEARCH CENTER

20 David W. Grissmer, Sheila Nataraj Kirby, Mark Berends, and Stephanie Williamson, Student Achievement and the Changing American Family (Santa Monica, CA: RAND, 1994).

21 Richard Rothstein with Karen Hawley Mills, Where's the Money Gone? Changes in the Level of Education Spending (Washington, D.C.: Economic Policy Institute, 1995), pp. 1 and 37. For an update for the 1991–96 period which shows a stagnation in spending on regular education and continued increases in special education spending, see Richard Rothstein, Where's the Money Going? Changes in the Level and Composition of Education Spending, 1991–96 (Washington, D.C.: Economic Policy Institute, 1997).

22 John E. Chubb and Terry Moe, Politics, Markets, and America's Schools, (Washington, D.C.: Brookings Institution, 1990).

23 Chubb and Moe's work has drawn strong support and considerable criticism. In their 1995 book, The Case Against School Choice, Kevin J. Smith and Kenneth J. Meier analyze the theoretical claims made, the methodology used, the results reported, and the conclusions drawn by Chubb and Moe. In addition they review data about the performance of choice programs in other countries. Smith and Meier conclude that the available empirical evidence did not support Chubb and Moe's case for vouchers. See Kevin B. Smith and Kenneth J. Meier, The Case Against School Choice: Politics, Markets, and Fools (Armonk, N.Y.: M.E. Sharp, 1995)

24 Stephen Herzenberg and Howard Wial, The State of Working Pennsylvania 1997 (Harrisburg, PA: Keystone Research Center, 1997).

25 166 Wis. 2d, 501, 480 N.W.2d, 460 (1992)

26 John F. Witte, First-Year Report: Milwaukee Parental Choice Program (Madison, WI: The Robert M. La Follette Institute of Public Affairs, University of Wisconsin-Madison, 1991); John F. Witte, Andrea B. Bailey, and Christopher A. Thorn, Second-Year Report: Milwaukee Parental Choice Program (Madison, WI: The Robert M. La Follette Institute of Public Affairs, University of Wisconsin-Madison, 1992); John F. Witte, Andrea B. Bailey, and Christopher A. Thorn, Third-Year Report: Milwaukee Parental Choice Program (Madison, WI.: The Robert La Follette Institute of Public Affairs, University of Wisconsin-Madison, 1993); John F. Witte, Christopher A. Thorn, Kim M. Pritchard, and Michelle Clairborn, Fourth-Year Report: Milwaukee Parental Choice Program (Madison, WI: The Robert La Follette Institute of Public Affairs, University of Wisconsin-Madison, 1994); John F. Witte, Troy D. Sterr, Christopher A. Thorn, Fifth-Year Report Milwaukee Parental Choice Program, (Madison, WI: The Robert LaFollette Institute of Public Affairs, University of Wisconsin-Madison, December 1995).

27 Paul E. Peterson, "A Critique of the Witte Evaluation of Milwaukee's School Choice Program," — Occasional Paper 95-2 (Harvard University Center for American Political Studies, 1995).

28 George A. Mitchell, The Milwaukee Parental Choice Program (Milwaukee: Wisconsin Policy Research Institute 1992); John E. Chubb and Terry M. Moe, Educational Choice: Answers to the Most Frequently Asked Questions About Mediocrity in American Education And What Can Be Done About It (Milwaukee: Wisconsin Policy Research Institute 1989).

29 Milwaukee Parental Choice Program, Wisconsin Legislative Audit Bureau Audit Summary Report 95-3 (Madison, WI.: Wisconsin Legislative Audit Bureau, 1995).

30 Witte, First-Year Report.; Witte, Bailey, and Thorn, Second-Year Report; Witte, Bailey, and Thorn, Third-Year Report; Witte, Thorn, Pritchard, and Clairborn, Fourth-Year Report; Witte, Sterr, and Thorn, Fifth-Year Report.

31 Witte, Sterr, and Thorn, Fifth-Year Report.

32 John F. Witte, "Achievement Effects of The Milwaukee Voucher Program," paper presented at the 1997 American Economics Association Annual Meeting, New Orleans, January 4-6, 1997.

33 Jay P. Greene, Paul E. Peterson, and Jiangtao Du with Jeesa Boeger and Curtis L. Frazier, "The Effectiveness of School Choice in Milwaukee: A Secondary Analysis of Data from the Program's Evaluation," August 14, 1996, Occasional Paper, Program in Education Policy and Governance, Center for American Political Studies, Department of Government, Harvard University; Jay P. Greene, Paul E. Peterson, and Jiangtao Du, "Effectiveness of School Choice: The Milwaukee Experiment," Occasional Paper 97-1, Program in Education Policy and Governance, Center for American Political Studies, Department of Government, Harvard University March,1997.



BEST COPY AVAILABLE

34 Rouse, "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program," Quarterly Journal of Economics, forthcoming.

35 Milwaukee Parental Choice Program, Wisconsin Legislative Audit Bureau Audit Summary Report 95-3.

36 Milwaukee Parental Choice Program, Wisconsin Legislative Audit Bureau Audit Summary Report 95-3.

37 The value of a MPCP voucher went up largely because the state has, since 1995, assumed a larger share of local education costs, leading to a large increase in per-pupil state aid to Milwaukee.

38 Ken Black, Milwaukee Public Schools Senior Budget Analyst, telephone interview December 19, 1997.

39 Brent Staples, "Schoolyard Brawl: The New Politics of Education Casts Blacks in a Starring Role," The New York Times, Section 4A (Education Life), January 4, 1998, p. 49.

40 Greene, Peterson, and Du, "The Effectiveness of School Choice in Milwaukee: A Secondary Analysis of Data from the Program's Evaluation," p. 6.

41 Witte, Thorn, Pritchard, and Clairborn, Fourth-Year Report.

42 William N. Evans and Robert M. Schwab, "Finishing High School and Starting College: Do Catholic Schools Make a Difference?" Quarterly Journal of Economics 110 (November 1995): 941-74. See also James S. Coleman, Thomas Hoffer, and Sally Kilgore, High School Achievement: Public, Catholic, and Private Schools Compared. (New York: Basic Books, 1982.)

43 Dan D. Goldhaber, "Public and Private High School: Is School Choice an Answer to the Productivity Problem?" Economics of Education Review, pp. 93-109.

44 David N. Figlio and Joe A. Stone, "School Choice and Student Performance: Are Private Schools Really Better?" Madison: Institute for Research on Poverty, Discussion Paper 1141-97, September 1997.

45 Cecilia Elena Rouse, "Schools and Student Achievement: More Evidence from the Milwaukee Parental Choice Program," Princeton and NBER, December 1997 (forthcoming, The Economic Policy Review.)

46 As the text indicates, each of the three research teams does other comparisons besides the one listed, in part because they have sometimes borrowed each other's methods as the debate over vouchers has proceeded.

47 Rouse, "Schools and Student Achievement: More Evidence from the Milwaukee Parental Choice Program," p. 11.

48 Witte, Sterr, and Thorn, Fifth-Year Report.

49 Greene, Peterson, and Du, "The Effectiveness of School Choice in Milwaukee: A Secondary Analysis of Data from the Program's Evaluation," p. 15.

50 In addition to critiquing Witte's regression analysis for the reason we give in the next, GPD attacked a wide variety of aspects of Witte's reports. Rather than reviewing each point, the text of the present paper emphasizes the key issues remaining after the dust had settled--and after the researchers had modified their approaches in response to criticism. For the caustic back and forth between Greene, Peterson, and Du and Witte, see, in addition to sources already cited: Paul E. Peterson, "A Critique of the Witte Evaluation of Milwaukee's School Choice Program," Center for American Political Studies Occasional Paper 95-2, Department of Government, Harvard University, 1995; John F. Witte, "A Reply to Paul Peterson's 'A Critique of the Witte Evaluation of Milwaukee's School Choice Program,'" University of Wisconsin, Department of Political Science, February 10, 1995; Paul E. Peterson, "The Milwaukee School Choice Plan: Ten Comments on the Witte Reply," Center for American Political Studies Occasional Paper 95-3, Department of Government, Harvard University, March 1995; John F. Witte, "Reply to Greene, Peterson, and Du: 'The Effectiveness of School Choice in Milwaukee: A Secondary Analysis of Data from the Program's Evaluation,'" University of Wisconsin, Department of Political Science, August 23, 1996; and Jay P. Greene and Paul E. Peterson, "Methodological Issues in Evaluation Research: The Milwaukee School Choice Plan," paper prepared for the Program in Education Policy and Governance, Department of Government and Kennedy School of Government, Harvard University, August 29, 1996.



BEST COPY AVAILABLE

51 Greene and Peterson, "Methodological Issues in Evaluation Research," p. 3 and Table 1.

52 Witte, "Achievement Effects of The Milwaukee Voucher Program"; and Rouse, "Private School Vouchers and Student Achievement," pp. 10-11.

53 Bruce Fuller, "It's Still Too Early to Judge School Vouchers," Letter to the editor The New York Times, May 17, 1997.

54 Greene, Paul E. Peterson, and Jiangtao Du, "Effectiveness of School Choice: The Milwaukee Experiment," Table 3. See also Greene, Peterson, and Du, "The Effectiveness of School Choice in Milwaukee: A Secondary Analysis of Data from the Program's Evaluation," p. 8 and, p. 19, table 4.

55 Witte, "Achievement Effects of The Milwaukee Voucher Program," In 1994, Witte had also done his own analysis of students who applied unsuccessfully to Choice schools and found no significant differences in the performance of Choice students and unsuccessful applicants. Witte, Thorn, Pritchard, and Clairborn, Fourth-Year Report, p. 25.

56 Rouse, "Private School Vouchers and Student Achievement," p. 23.

57 Rouse gives an unusually clear explanation of her fixed effects approach in Cecilia Elena Rouse, "Schools and Student Achievement," pp. 8-9 and Figure 1.

58 For her discussion of these, see especially Rouse, "Private School Vouchers and Student Achievement," p. 33 and p. 19.

59 Rouse, "Schools and Student Achievement," p. 17.

60 Terry Moe, ed., Private Vouchers, (Stanford, CA: Hoover Institution Press, 1995), p. 45.

61 Marie Rohde, "Minority Test Scores at Catholic Schools Mirror Lag in City," Milwaukee Journal, August 1, 1991.

62 Ernst-Ulrich Franzen, "Archdiocese Abolishes School System," Milwaukee Sentinel, January 27, 1994.

63 Milwaukee Public Schools Governmental Relations Office, correspondence to the author, December 1 and 6, 1995.

64 Emily Koczela, Timothy J. McElhatton, and Jean B. Tyler, Public and Private School Costs: A Local Analysis (Milwaukee: Public Policy Forum, 1994).

65 Janet R. Beales and Maureen Wahl, "Private Vouchers in Milwaukee: The PAVE Program," in Terry Moe, ed., Private Vouchers, Stanford, California: Hoover Institution Press, 1995.

66 Maureen Wahl, First-year Report of the Partners Advancing Values in Education Scholarship Program (Milwaukee: Family Service America, 1993); Maureen Wahl, Second-year Report of the Partners Advancing Values in Education Scholarship Program (Milwaukee: Family Service America, 1994); Maureen Wahl, Third-year Report of the Partners Advancing Values in Education Scholarship Program (Milwaukee: Family Service America, 1995)

67 Sammis B. White, Peter Maier, and Christine Cramer, Fourth-Year Report of the Partners Advancing Values in Education Scholarship Program, (Milwaukee: Urban Research Center, University of Wisconsin-Milwaukee, 1996)

68 The description of the Cleveland Scholarship and Tutoring Program is based on documents provided by the Ohio Department of Educaton, discussions with Francis Rogers of the Ohio Department of Education, and Bert Holt of the Cleveland Scholarship and Tutoring Program, and Dan Murphy, F. Howard Nelson, Bella Rosenberg, The Cleveland Voucher Program: Who Chooses? Who Gets Chosen? Who Pays?, American Federation of Teachers, 1997.

69 Murphy, Nelson, and Rosenberg, The Cleveland Voucher Program: Who Chooses? Who Gets Chosen? Who Pays?

70 Jay P. Greene, William G. Howell, Paul E. Peterson, An Evaluation of the Cleveland Scholarship Program, Harvard University, Program on Education Policy and Governance, September 1997.

71 For a discussion of the problems posed by fall-to-spring testing in Chapter I programs see: Robert L. Linn, Stephen B. Dunbar, Delwyn L. Harnish, and C. Nicholas Hastings, "The Validity of the Title I Evaluation and Reporting System," in Ernest R. House, Sandra Mathison, James A. Pearsol, Hallie Preskill, ed., Evaluation Studies Review Annual, 7 (Beverly Hills: Sage Publications, 1982).

72 P.W. Cookson, School Choice: The Struggle for the Soul of American Education (New Haven: Yale University Press, 1994)

73 Henig, Rethinking School Choice.

74 Geoffrey Schneider, Bucknell University Department of Economics, unpublished notes and conversations with Keystone Research Center staff. The argument given in the text draws on Schneider's analysis but is not identical to it.

75 Moe, ed., Private Vouchers.

76 Moe, ed., Private Vouchers.

77 School Choice (Princeton: The Carnegie Foundation for the Advancement of Teaching, 1992).

78 Bruce Fuller, "Who Gains, Who Loses from School Choice: A Research Summary," policy brie, National Conference of State Legislatures, Denver: 1995.

79 Geoff Whitty, "Creating Quasi-Markets in Education: A Review of Recent Research on Parental Choice and School Autonomy in Three Countries," in Michael W. Apple, ed., Review of Research in Education, 22 (Washington, D.C.: American Educational Research Association, 1997)

80 Martin Carnoy, "Is School Privatization the Answer? Data from the Experience of Other Countries Suggest Not." Education Week, July 12, 1995.

81 Eric A. Hanushek, "The Impact of Differential Expenditures on School Performance," Educational Researcher, 18, (May 1989): 45-65.

82 Larry V. Hedges, Richard D. Laine, and Rob Greenwald, "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes," Educational Researcher, 2,3 (April 1994): 5-14. For the continuing back and forth between Hanushek and his critics, see Eric A. Hanushek, "School Resources and Student Performance," in Gary A. Burtless (ed.). Does Money Matter? (Washington, D.C.: Brookings Institution, 1996); Erik A Hanushek, "Assessing the Effects of School Resources on Student Performance; An Update," Educational Evaluation and Policy Analysis, 1962 (1997): 141–164; and Robert Greenwald, Larry V. Hedges, and Richard D. Laire, "The Effect of School Resources on Student Achievement," Review of Educational Research, 66 (3) (1996): 391–396

83 Bruce J. Biddle, "Foolishness, Dangerous Nonsense, and Real Correlates of State Differences in Achievement," Phi Delta Kappan, 79 (September 1997): 9-13.

84 Rothstein Where's the Money Gone? Changes in the Level of Education Spending.

85 Bruce J. Biddle, "Foolishness, Dangerous Nonsense, and Real Correlates of State Differences in Achievement," pp. 9-13.

86 Ronald F. Ferguson, "Paying For Public Education: New Evidence On How and Why Money Matters," Harvard Journal on Legislation, 28: 465-498.

87 Harold Wenglinsky, "How Money Matters: The Effect of School District Spending on Academic Achievement," Sociology of Education, (July 1997): 221-237.

88 David N. Figlio and Joe A. Stone, "School Choice and Student Performance," p. 15.

KEYSTONE RESEARCH CENTER

**Small Class Sizes**

1 Glen E. Robinson and James H. Wittebols, "Class Size Research: A Related Cluster Analysis for Decision Making," ERS Research Brief ED No. 274-030 (Arlington, VA: Educational Research Service, Inc., 1986).

2 Douglas Mitchell, Cristi Carson, Gary Badarak, How Changing Class Size Affects Classrooms and Students, (Riverside, CA: California Educational Research Cooperative, School of Education, University of California-Riverside, 1989).

3 Robinson and Wittebols, "Class Size Research."

4 Mitchell, Carson, Badarak, How Changing Class Size Affects Classrooms and Students.

5 Gene V. Glass and Mary Lee Smith, Meta-Analysis of Research on The Relationship of Class-Size and Achievement (San Francisco, CA: Far West Laboratory for Educational Research and Development, 1978). See also Gene V. Smith, Leonard S. Cahen, Mary Lee Smith, and Nikola N. Filby, School Class Size: Research and Policy (Beverly Hills, CA: Sage Publications, 1982).

6 Robinson and Wittebols, "Class Size Research."

7 Robert E. Slavin, "Meta-analysis in Education: How Has It Been Used?" Educational Researcher, 24, 1984.

8 National Center for Education Statistics, Digest of Education Statistics (Washington, D.C.: U.S. Government Printing Office, forthcoming 1997), Table 64, page number not yet assigned.

9 National Center for Education Statistics, Schools and Staffing in the United States: A Statistical Profile 1993-94, (Washington, D.C.: U.S. Government Printing Office, 1996), Table A15, p. 166.

10 National Center for Education Statistics, Digest of Education Statistics 1996, (Washington, D.C.: U.S. Department of Education, 1996), Table 87, p. 95.

11 Michael Boozer and Cecilia Rouse, "Intraschool Variation in Class Size: Patterns and Implications," Princeton University, Industrial Relations Section, Working Paper #344, June 1995.

12 For a current synopsis of Prime Time, see Indiana Department of Education, "Prime Time: An Overview," no date, available from Mary Beth Morgan or Jayma Ferguson by calling (317) 232-9163.

13 David Gilman and Christopher Tillitski, "The Longitudinal Effects of Smaller Classes: Four Studies," October 1990, ED # 326 313

14 B. Farr et. al, Evaluation of Prime Time: 1986-87, Final Report (Indianapolis, IN: Advanced Technology, no date); Mary Quilling, Linda Parker, and David Gray, Prime Time: Six Years Later, prepared for the Indiana Department of Education by PRC, May 1992.

15 Quilling, Parker, and Gray, Prime Time: Six Years Later, p. 20.

16 Quilling, Parker, and Gray, Prime Time: Six Years Later, p. 25.

17 Quilling, Parker, and Gray, Prime Time: Six Years Later, p. 29.

18 H. Bain, C.M. Achilles, B. Dennis, M. Parks, and R. Hooper, "Class Size Reduction in Metro Nashville: A Three Year Cohort Study," ERS Spectrum, 6(4) (1988): 30-36.

19 Elizabeth Word, John Johnstone, Helen Pate Bain, B.DeWayne Fulton, Jayne Boyd Zaharias, Martha Nannette Lintz, Charles Achilles, John Folger, Carolyn Breda, Student/Teacher Achievement Ratio (STAR) Tennessee's K-3 Class Size Study, Final Summary Report 1985-1990 (Nashville, TN: Tennessee Department of Education, 1990).

20 Alan B. Krueger, "Experimental Estimates of Education Production Functions," Princeton University Industrial Relations Section, Working Paper #379, May 1997.

21 Alan B. Krueger, "Experimental Estimates of Education Production Functions."

22 Jayne Boyd-Zaharias, C.M. Achilles, Barbara A. Nye, Helen Pate Bain, B.D. Fulton, "Quality Schools Build On A Quality Start," in Edward W. Chance, ed., Creating the Quality School (Madison, WI: Magna Publications, 1995). In regression analysis, Alan Krueger also considers the impact of using actual class size rather than the original designations into small and regular classes. As in the Boyd-Zaharias study, having a class size difference of more than 10 generates a significantly larger effect size than when considering the originally-designated small and regular classes as a group (Alan B. Krueger, "Experimental Estimates of Education Production Functions.")

23 Word, et al., Student/Teacher Achievement Ratio(STAR)Tennessee's K-3 Class Size Study, Final Summary Report 1985-1990. These findings are also reported in Jeremy D. Finn, Charles M. Achilles, Helen Pate Bain, John Folger, John M. Johnston, M. Nan Lintz, Elizabeth R. Word, "Three Years In A Small Class," Teaching and Teacher Education, 6(2), 1990: 127-136.

24 Jeremy D. Finn and Charles M. Achilles, "Answers and Questions About Class Size: A Statewide Experiment," American Educational Research Journal, 27(3) (Fall 1990): 557-577.

25 Alan B. Krueger, "Experimental Estimates of Education Production Functions."

26 Charles M. Achilles, Jeremy D. Finn, and Helen P. Bain, "Using Class Size to Reduce the Equity Gap," Educational Leadership, 55 (4) (December 1997/January 1998): 40-43.

27 C. Steven Bingham, White-Minority Achievement Gap Reduction and Small Class Size: A Research and Literature Review (Nashville, TN: Center of Excellence for Research and Policy on Basic Skills, Tennessee State University, 1993).

28 John Folger and Carolyn Breda, "Evidence From Project STAR About Class Size and Student Achievement," Peabody Journal of Education, 67(1) (Fall 1989): 17-33.

29 J. Fowler, Jr., and J. Walberg, "School Size, Characteristics and Outcomes," Educational Evaluation and Policy Analysis, 13(2) (Summer 1991): 189-202.

30 C.M. Achilles, Summary of Recent Class-Size Research With an Emphasis On Tennessee's Project STAR and Its Derivative Research Studies (Nashville, TN: Center of Excellence for Research and Policy on Basic Skills, no date).

31 Frederick Mosteller, "Tennessee Study of Class Size in the Early School Grades," The Future of Children, 5(2), (Summer/Fall 1995).

32 The LBS (1992-to date) follow-up to STAR (1985-89) led to the establishment of a full scale state-wide K-3 class size reduction policy in 1995. Under this policy, Tennessee will provide state funding to reduce state-wide K-3 class size to 20:1 with an option of 15:1 for at-risk students. These goals should be achieved by the year 2001 and are already 80 percent achieved in kindergarten and grade one. The STAR and LBS projects and subsequent legislative action are an excellent example of the effective use of research and evaluation by state policy makers.

33 In the eight-grade technical report, about half the sample had been in small K-3 classes. About 20 percent of the sample had only one year in a small class. Barbara A. Nye, B. DeWayne Fulton, Jayne Boyd-Zaharias, Van A. Cain, The Lasting Benefits Study, Eighth Grade Technical Report (Nashville, TN: Center of Excellence for Research in Basic Skills, Tennessee State University, 1995), Table 1, p. 5.

34 Nye, Fulton, Boyd-Zaharias, and Cain, The Lasting Benefits Study, Eighth Grade Technical Report. See also Barbara A. Nye, "Class Size and School Effectiveness: Answers from Longitudinal Studies," paper presented at the 11th International Congress for School Effectiveness and School Improvement, University of Manchester, U.K., January 1998.

35 Nye, Fulton, Boyd-Zaharias, and Cain, The Lasting Benefits Study, Eighth Grade Technical Report, Table 5 and Table 7.

36 Barbara A. Nye, "Class Size and School Effectiveness."

37 Barbara A. Nye, Jayne Boyd-Zaharias, B. DeWayne Fulton, C.M. Achilles, "The Lasting Benefits Study: Seventh Grade Technical Report" (Nashville, TN: Center of Excellence for Research in Basic Skills, Tennessee State University, 1994).

38 Barbara A. Nye, Jayne Boyd-Zaharias, B. DeWayne Fulton, C.M. Achilles, The Lasting Benefits Study, Sixth Grade Technical Report (Nashville, TN: Center of Excellence for Research in Basic Skills, Tennessee State University, 1993).

39 Barbara A. Nye, B. DeWayne Fulton, and Jayne Boyd-Zaharias,The Lasting Benefits Study, Fifth Grade Technical Report (Nashville, TN: Center of Excellence for Research in Basic Skills, Tennessee State University, 1992.)

40 Barbara A. Nye, Jayne Boyd-Zaharias, B. DeWayne Fulton, C.M. Achilles, The Lasting Benefits Study, Fourth Grade Technical Report (Nashville, TN: Center of Excellence for Research in Basic Skills, Tennessee State University, 1991).

41 Barbara A. Nye, Charles M. Achilles, Jayne Boyd-Zaharias, and B. DeWayne Fulton, Project Challenge Fifth Year Summary Report (Nashville, TN: Center of Excellence for Research in Basic Skills, College of Education, Tennessee State University, 1995). See also the first (1991), second (1992), and third (1993) Project Challenge summary reports and Charles M. Achilles, Barbara A. Nye, Jayne Boyd-Zaharias, and B. DeWayne Fulton, "The Lasting Benefits Study (LBS) In Grades 4 and 5 (1990-1991): A Legacy From Tennessee's Four-Year (K-3) Class-Size Study (1985-1989), Project STAR," paper presented at the North Carolina Association for Research in Education, Greensboro, North Carolina, January 14, 1993.

42 Nye, Fulton, Boyd-Zaharias, and Cain, Project Challenge Fourth Year Summary Report, p. 9; Frederick Mosteller, "The Tennessee Study of Class Size In The Early School Grades," pp. 122-123. See also Charles M. Achilles, Barbara A. Nye, Jayne Boyd-Zaharias, "Policy Use of Research Results: Tennessee's Project Challenge," paper presented at the American Educational Research Association, San Francisco, CA, April 1995.

43 For data on class size over time in Nevada in grades K-3, see H. Pepper Sturm, "Nevada's Class-size Reduction Program," Senate Human Resources Committee, Nevada Legislative Counsel Bureau, February 1997.

44 Telephone interview, January 14, 1998.

45 Mary Snow, "An Evaluation of the Class Size Reduction Program," Carson City, NV: Nevada Department of Education, May 1997, p. 5.

46 Mary Snow, "An Evaluation of the Class Size Reduction Program," p. 3.

47 James Pollard and Kim Yap, "The Nevada Class Size Reduction Evaluation Study, 1995," Carson City, NV: Nevada State Department of Education, 1995, p. 10.

48 Judith S. Costa and Rhoton Hudson, "Report Prepared by Testing and Evaluation Department," Clark County School District, February 1995, p. 5.

49 This section relies primarily on Joan McRobbie, "Class Size Reduction: Is It Working?" Thrust for Educational Leadership, September 1997.

50 Professor Michael Kirst, Stanford University, telephone interview, January 16, 1998.

51 Diane Flores, "Class Size Survey," Modesto Bee, June 20, 1997 (available via the San Diego Count Office of Education Web Page at http://www.sdcoe.k12.ca.us/csr_survey/bee.html).

52 Available via the San Diego County Office of Education Web Page at http://www.sdcoe.k12.ca.us/csr_survey/. For the results of a third survey, in Santa Barbara County, which found positive parent and teacher perceptions of the class size initiative, see Bill Cirone, "Survey Finds Overwhelming Support for Class Size Reduction," Thrust for Educational Leadership, September 1997.

53 As this report went to press, a draft "Short-term Class-Size Reduction Study" which pulls together all findings collected to data was not yet public. The study does not contain any information on test scores. Telephone interview with Joan McRobbie, January 16, 1997.

54 A consortium comprehensive evaluation plan for class size reduction has been submitted to the California State Board of Education but is not a public document.

55 Rouse, "Schools and Student Achievement."

56 The term pupil-teacher ratio is used here because of number of participating schools reduced their class sizes by means other than placing one teacher with 15 students. In the SAGE program, "pupil-teacher ratio" is used in ways that capture the everyday understanding of "class size." It is not a statistical artifact, for example, of having many certified staff outside the classroom.

57 Maier, Alex Molnar, Philip Smith, John Zahorik, First-year Results of the Student Achievement Guarantee in Education Program (Milwaukee: Center for Urban Initiatives and Research, University of Wisconsin-Milwaukee, December 1997).

**Recommendations**

1 Harold Wenglinsky, When Money Matters (Princeton, NJ: Educational Testing Service), pp. 23-24. This source (p. 24) shows an average gain for math students in smaller-than-average classes of about half-a-year compared to students in larger-than-average classes. In a more recent paper, Wenglinsky uses a more sophisticated approach to control for differences in socio-economic status and other variables across schools. When he does this, he finds that a class of 15 students would be four months ahead of a class of 25 (telephone interview, January 27, 1998, and Harold Wenglinsky, Modeling the Production Function: Associates Among School Districts Expenditures, School Resources and Student Achievement (Princeton, NJ: Educational Testing Service, November 1997).)

2 Alan B. Krueger, "Experimental Estimates of Education Production Functions," p. 27.

3 See, for example: John Folger, "Lessons for Class Size Policy and Research," Peabody Journal of Education, 67(1) (Fall 1989): 122-132; John Folger and Carolyn Breda, "Evidence From Project Star About Class Size an Student Achievement," Peabody Journal of Education, 67(1) (Fall 1989): 17-33; Allan Odden, "Class Size and Student Achievement: Research-Based Policy Alternatives," Educational Evaluation and Policy Analysis, 12(2) (Summer 1990): 213-227; Robert E. Slavin, in Robert E. Slavin, Nancy L. Karweit, and Barbara A. Wasik, ed., "School and Classroom Organization in Beginning Reading: Class Size, Aides, and Instructional Grouping," Preventing Early School Failure: Research, Policy; and Practice (Boston, MA: Allyn and Bacon).

4 In 1996-97, Pennsylvania had about 134,000 children in kindergarten and 148,000 in first grade (according to the Pennsylvania Department of Education). Targeting class-size reduction at the quarter of Pennsylvania children most in need (including those of kindergarten age who do not attend public school) would impact about 37,000 students in kindergarten and another 37,000 the next year in first grade. In kindergarten, roughly one half of these 37,000 children are in half-day kindergarten (state-wide, only 26 percent of children are in full-day kindergarten; but a larger share of low-income students are in full-day kindergarten, including almost all Philadelphia School District students and a majority of Pittsburgh schoolchildren). If one assumes that classes enroll 22.5 students each today, then 37,000 students split between half and full-day kindergarten requires 1233 kindergarten teachers. Providing full-day kindergarten with a student-teacher ratio of 1:15 requires 2467 teachers, an increase of 1234. If new teachers are paid about $35,000-$40,000 per year, including benefits, the extra 1234 teachers would cost $43-49 million dollars in year one. Since first grade is already all day everywhere, the number of teachers to go from a 22.5 ratio to 15 would simply be 50 percent more than the current number of first grade teachers. This is an increase of 822 teachers, at a cost of about $29-$33 million (note that this range is similar to the $29.6 million that California would allocate for 37,000 students at its current funding level of $800 per student). In sum, over two years, the projected cost is about $115 million in teacher compensation. $100 million per year should therefore be adequate to cover facilities, other ancillary costs, and a comprehensive research evaluation, as well as teacher salaries.

BEST COPY AVAILABLE

BEST COPY AVAILABLE

For additional copies of this report contact:

KEYSTONE
RESEARCH
CENTER

Keystone Research Center • 412 North Third Street • Harrisburg, PA 17101
Phone: 717-255-7181 • Fax: 717-255-7193 • Internet: KeystoneRC@aol.com